

UNITEXT for Physics

Giuseppe Iadonisi
Giovanni Cantele
Maria Luisa Chiofalo

Introduction to Solid State Physics and Crystalline Nanostructures

 Springer

UNITEXT for Physics

Series editors

M. Cini, Roma, Italy
A. Ferrari, Torino, Italy
S. Forte, Milano, Italy
I. Massimo, Firenze, Italy
G. Montagna, Pavia, Italy
O. Nicrosini, Pavia, Italy
L. Peliti, Napoli, Italy
A. Rotondi, Pavia, Italy

For further volumes:

<http://www.springer.com/series/13351>

Giuseppe Iadonisi • Giovanni Cantele
Maria Luisa Chiofalo

Introduction to Solid State Physics and Crystalline Nanostructures



Springer

Giuseppe Iadonisi
University of Napoli “Federico II”
Italy

Giovanni Cantele
SPIN Institute, National Research Council
Italy

Maria Luisa Chiofalo
University of Pisa and INFN
Italy

ISSN 2198-7882
ISBN 978-88-470-2804-3
DOI 10.1007/978-88-470-2805-0
Springer Milan Heidelberg New York Dordrecht London

ISSN 2198-7890 (electronic)
ISBN 978-88-470-2805-0 (eBook)

Library of Congress Control Number: 2013955592

© Springer-Verlag Italia 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Franco Bassani, role-model teacher and beloved friend

Preface

Why this textbook

Who would not like to be in several places at the same time, tunnel through barriers, send out information or items without real traveling, or else run a computer one million times faster than home laptop, for example able to perform a web or a database search some billion times faster than present computers. Among those who dream about skies and stars, who would not like to know more and better about how the Universe works, being able to orient him or herself among those ideas that are most useful to understand what goes on at the tiniest length scales and enormous energies below the nuclei constituents, where quantum physics and relativity dominate, and being able to link these ideas to what goes on instead at the giant cosmic distances where gravity and relativity do dominate. What appears impossible in our macroscopic world is instead possible in microscopic, quantum world. To this aim, a progressive search for realizing and investigating extreme quantum-degeneracy conditions has been conducted in the last few decades: along these lines, the physics of solid state and of nanostructures has been a rapidly evolving field of frontier research, exploding in a number of novel ideas with applications for both fundamental physics and technological devices.

Extreme quantum conditions are reached whenever the length scale of typical de Broglie wavelengths of matter and of system sizes are comparable. This might occur for example after reducing the dimensions from three down to zero-dimensional like structures, lowering temperatures so close down to absolute zero and therefore heavily reducing the average kinetic energies, tuning interaction energies to large values via a variety of different concepts, so to make kinetic energy comparably less significant. Novel lowest-energy, i.e. ground states of matter and related low-energy types of excitations have been predicted and in most cases experimentally realized: it is often possible to move among different ground states in the same quantum phase diagram by turning the knob of a few external governing parameters. Engineering such extreme quantum conditions has opened the access to a progressive miniaturization, control and manipulation of quantum-behaving objects with high precision. These characteristics are those required to improve the capability of understanding

and testing fundamental physics concepts, such as those principles that cross over from quantum behavior to general relativity, and it is required to devise novel technological applications, such as those based on quantum transport and quantum information. Laboratory systems where these frontier research topics are quite rapidly evolving are for example semiconductor-based nanostructures and, more recently, confined quantum atomic gases cooled down to tens of nanokelvin temperatures, moving within crystal-like periodic external potentials.

The progress achieved so far has been accompanied by significant advances in experimental and theoretical methods. New concepts have indeed been introduced in experimental frameworks, such as various specifications of probe microscopies, and earlier spectroscopic techniques have been refined to reach unprecedented accuracies and precision. Powerful theoretical methods specific for many-particle physics have been developed, where for example various specifications of zero and finite-temperature simulational Quantum Monte Carlo methods and of Density Functional Theories have occupied a preeminent position. More recently, further increase in computation powers and design of clever algorithms has made available exact diagonalization numerical techniques to determine ground states and low-energy excitations and Density Matrix Renormalization Group techniques to determine the time-dependent behavior of even strongly correlated systems.

With the considerations above, it is of no surprise that at such advanced levels and especially in regard to semiconductor nanostructures, physics, chemistry and engineering are converging towards the use of common tools and languages though from different motivations and with different perspectives.

A number of excellent solid-state physics textbooks for undergraduate studies have been published in the last fifty years, that have represented the food for thought and training of generations of physicists. Some among them require the knowledge and skillful use of advanced mathematical tools. Some others, for beginners, are shaped and focused on the needs of either physics, or chemistry, or engineering students. We have therefore been moved by the idea of making available a textbook that be at the same time interesting for physicists, chemists, and engineers, while asking for basic knowledge of mechanics, thermodynamics, electromagnetism, and quantum mechanics as prerequisites. A strategy pointing towards a multidisciplinary approach may favor the acquisition of cross competences, reinforce the attitude towards the scientific method from different paths, widen the student's view while the same ideas are discussed from different perspectives and, paradoxically, push forward the emergence of the key-concepts as those that are shared across all the different viewpoints. Finally, it may increase the qualitative and quantitative impact of understanding along with the activation of different languages and thus communication channels that can more likely meet different students abilities.

In fact, physics is an extraordinary context for learning, often at a deeper level than other disciplines. In physics, the learning success at any level, even the lowest, strongly requires the development of different abilities, that become competences when accompanied by the corresponding awareness, and evolve into intelligences when put in context. Different intelligences residing into different brain sites and functions. The ability to solve problems, that is to reduce the complexities of na-

ture and to select the useful amount and quality of details, requires the acquisition of robust methodological paths able to connect concepts and ideas with their applications, thereby linking conceptual with procedural and factual knowledge. This crossing from brain to hands and back, from concepts to their applications, from the observation of phenomena to inferring explanations that are to be subsequently tested, works as a continuous exercise of climbing up and down across all types of logical reasoning, whatever deductive, inductive or abductive they might be. The effort to reduce the great amount of many different behaviors to a few basic principles requires an active lateral thinking and a remarkable dose of creativity. The ability to continuously change perspective in flexible manners requires the acquisition of a number of different languages or, more precisely, formal representations of the problem, along with the ability of translating among them and often of thinking according to all of them at the same time. On the other hand, since the capability of quantitatively predicting complex phenomena requires significant mathematical tools and physics formalisms, students and future researchers can be in danger of remaining trapped in a crowd of hard calculations and technicalities where the essential concepts get lost as a compass to orient the understanding in the widest horizon, and the contact with everyday life manifestations of phenomena is not even an issue. For this reason, a special attention is to be dedicated in discussing ideas and concepts, well before entering any formalism and detail, which makes the learning process additionally powerful.

The possibility of developing a high degree of awareness for this remarkable learning potential is a progressive long-life achievement for students and researchers, which the format of the present textbook tries to smoothly follow through. A special care has therefore been taken to think the textbook organization: the aim is to enhance the student's learning environment with suited tools to develop awareness of the learning process and of the acquired conceptual, procedural, and factual knowledge. In this way, the original combination of student's skills, different for everyone, may more easily become a competence and, when related and applied to a context, may empower a different combination of intelligences.

What this textbook is about

According to the motivations discussed so far, this textbook is aimed to general and specific goals as follows.

General goals

1. Bring science to students, specifically solid-state physics, in a personalized way. Each student is different with her/his own abilities, learning styles, combination of intelligences via the skills in the use of different languages.
2. Favor the acquisition of scientific method.
3. Accompany students in the process of focusing the main concepts and connections between concepts, that are physics laws, and make physics understandable and interesting well before technicalities get in.
4. Empower students with the development of their intuition and creativity.

5. Accompany students in the process of acquiring strategies and procedures for the solution of typical problems, reinforcing the idea that solving a physics problem is everything but a magical matter reserved to especially talented persons.
6. Trigger students curiosity about how low-energy matter works and therefore empower their motivation to make and invent new solid-state physics.
7. Favor the process of awareness about what has been learned, thus pursuing the meta-goal of empowering the students skills for science communication as an effective competence to be exploited at any level: as researchers when working in a group or within a community, as fundraisers to explain the importance of a given research activity to the decision makers who address the resources, as teachers at the first school degrees, or as scientists committed to bring science to next-door neighbors.

Specific goals

1. The main specific goal is to provide the student with a training for essential concepts and methods in the field of solid-state physics, at a level depth that be sufficient to give an overview for the student who does not intend to specialize in the field, and to be an useful base for the student who is going to enter advanced solid-state studies. In any event, a selection of topics has of course been operated, so that the content is by no means exhaustive, not even at a surface level.
2. The focus on the essentials of nanostructures represent the first and most evident choice that has been operated. This is due to one content-like and one methodological-like motivation. As to the content, the physics, chemistry and engineering of semiconductor-based nanostructures has been one of those rapidly evolving fields of research for both fundamental physics and technological applications, where advances in experimental and theoretical methods have been mostly involved. Therefore, nanostructures offer a wide and effective view of the more recent progress in solid-state physics. From a methodological and educational perspective, nanostructures represent a natural system to bridge the properties of single atoms and molecules with those of macroscopic crystals. The evolution from nano- to macroscopic regimes and back is a smooth process in terms of the system size. This process can be followed while investigating and determining the behavior of physical quantities and therefore naturally works as a tutorial in building up the useful knowledge while proceeding from simpler to more complex and from known to unknown situations. Each physical quantity that is involved possesses a typical scale size, which the nanoscopic behavior begins to manifest below. In particular, quantum effects dominate as long as the structure is reduced to sizes comparable to the de Broglie wavelength of the particles. Electronic, optical, and transport properties of materials considerably depend on chemical composition, environmental parameters, short- or long-range structural ordering, the occurrence of quantum effects. A nanostructure can therefore be considered as an atomic or molecular aggregate with sizes ranging from tens to hundreds Angstrom, that is from the size of a large molecule up to that of a tiny crystal where the long-range ordering begins to show up. In order to catch the appropriate length scales, consider the representation of a crystal as a number of cubic structures with side a replicated back-to-back in the three-dimensional lattice, with one atom at each cube vertex. If n is the num-

ber of cubes per line, one would have $N_s = (6/a^2)(na)^2$ atoms in the surface and $N_t = a^{-3}(na)^3$ in the whole volume. The ratio $N_s/N_t = 6/n$ is a decreasing function of n , with $N_s/N_t \simeq 1$ as $n < 10$ and $N_s/N_t < 10^{-3}$ as $n > 6000$. One sample with 6000 cubic units would have a size $\simeq 10^{-5}$ cm, in the range of visible light wavelength and much larger than the typical lattice size, that is of the order of $\simeq 10^{-8}$ cm. Electronic and vibrational properties for example are characterized by macroscopic behavior when $N_s/N_t \ll 1$, and by nanoscopic behavior when $N_s/N_t \sim 1$. Transport behavior is instead driven by the average path λ that charge carriers walk between each two scattering events with lattice ions, on the scale of the sample size L : the condition $\lambda/L \ll 1$ signals macroscopic behavior whereas $\lambda/L \geq 1$ the nanoscopic one. In general, a sample can show up nanoscopic behavior for electronic or vibrational properties and macroscopic behavior for transport properties and viceversa, as the two conditions on N_s/N_t and λ/L are not necessarily overlapping to each other.

3. An additional selection concerns the choice of which other ground and excited states are usefully investigated within the textbook, different from the ordinary classification of solid-state matter in metals, semiconductors and insulators. Here we have decided to illustrate the essentials of superconducting and superfluid materials, as manifestations of at least two additional crucial concepts: pairing in the presence of many-particle correlations and its consequences for coherent transport, and the peculiar response of superfluids to shear external perturbations. In addition, we have included a basic description of magnetic orderings, such as they occur in ferro- and antiferromagnetic materials, as peculiar and all-quantum manifestations of particles spin. It is clear that these two selections are aimed to just make students aware of the fact that the complications arising from charge, density or spin-driven interactions, along with the possibility of tuning them in engineered manner, are what makes solid-state matter so interesting for applications and fundamental-physics tests. Many other peculiar and novel ground states and low-energy excitations exist and can be possibly conceived, for which the student is referred to more advanced textbooks. In all cases, whenever possible examples are discussed of the manifestation of these ground states in systems different from ordinary crystals, such as for example confined ultracold quantum gases in periodic potentials provided by laser light and mimicking a sort of crystal of light.

4. The last selection is related to the methodological approach used to systematically include the correlations between particles. In this respect, we have decided to give a special weight to Density Functional Theories (DFT), to which a full Chapter, the last one, is dedicated. The motivation is threefold. First, Density Functional Theory and its extensions to time-dependent, superfluid, and magnetic materials is a worth framework per se: DFT has been demonstrated to represent, conceptually and practically, a powerful tool to describe and predict the behavior of strongly correlated and inhomogeneous systems, often joined to Quantum Monte Carlo simulational methods or advanced perturbative expansion calculations. Second, the basics of Density Functional Theory can be caught with almost ordinary mathematical tools at hand. Third, Density Functional Theory and its extensions work extraordinarily well as a tutorial, smooth, introduction of correlations in terms of functionals of the density, starting with Hartree and Hartree-Fock approximations and ending up with the

exact formulation in terms of the Hohenberg and Kohn theorem and the practical implementation provided by the Kohn-Sham scheme.

How this textbook is organized

Chapters ordering. The order with which the whole matter is presented reflects the unsophisticated idea that crystals and nanostructures, at least at the present low meV - to -keV energies and nanoscopic length scales, are composed by: ions arranged in a lattice-like structure, which vibrate around the equilibrium lattice positions; electrons or missing electrons, i.e. holes or more generally charged moving particles; spins carried by these charged moving particles. In addition, these crystal constituents can be exposed to external perturbations that can be generally classified to be either mechanical or electromagnetic in nature. Thus, Chapter 1 describes the structure of crystals as it can be inferred from the static positions of ions arranged in the regular lattice. The conditions for crystal formation and the cohesive energy are discussed here, also anticipating the connections with various types of chemical bonding. Advantage is taken from the definition of crystal, also to provide a dictionary on amorphous and liquid materials, giving at least at conceptual level an illustration of the meaning of glasses, polymers and plastics. The description of the material behavior under mechanical strains is anticipated here and then formally developed at the beginning of Chapter 3. Chapter 2 builds up the language to describe and the tools to determine the structure of ground states in which electrons are allowed to live in and the excited states where they are allowed to move to. The classification of crystals in metals, semiconductors and insulators is given here, and at a less deeper level the ground and excited states of magnetically ordered and superconducting materials. Chapter 3 faces the problem of determining the structure of allowed vibrational excitations characterizing the motion of ions, that so far have been treated as fixed at the equilibrium lattice-sites positions. Up to this Chapter, the allowed ground and excited states of ions and charge and spin carriers have been treated. Chapter 4 introduces the tools to describe the behavior of charge and spin carriers, once an appropriate electric, magnetic, or temperature-gradient external field is applied: besides charge and spin, electrons, holes, and ions may in general be responsible for an energy flux. This transport behavior is characterized for metals, semiconductors, and superconductors. Chapter 5 introduces the tools to describe the system behavior once an external electromagnetic wave, notably light of given frequency, impinges on the sample. The resulting optical properties are investigated, to which both charge and vibrational excitations in general contribute to. Chapters 4- 5 share two key ideas. First, the response of the system to external time and space-dependent perturbations reflects the power dissipated by the system once a fluctuation created by the external disturbance washes away, that is the content of the fluctuation-dissipation theorem. Second, once the system characteristics under investigation are given, the external perturbation has to be chosen with appropriate type and symmetry, energy and wavelength, so to effectively couple to the corresponding system observable. These key concepts are first illustrated by means of the simple classical model of a forced, damped harmonic oscillator and then gener-

alized to more refined quantum situations. Up to this Chapter, repulsive or attractive interactions among electrons or holes, among spins, or among ions and electrons or holes, as well as the exchange effects due to the quantum statistics of the particles have been introduced in each Chapter at the lowest level of approximation, either to account for otherwise unexplainable experimental observations or to provide a general idea of how the whole picture would be modified. Most often, the interactions are seen to be expressed in terms of functionals of the density, though within an approximated approach. This simple idea is refined and extended in the final Chapter 6, based on the Hohenberg and Kohn theorem and the Kohn-Sham scheme. Here, the general framework of Density Functional Theory is set, providing a powerful conceptual and practical tool to treat a strongly correlated slightly inhomogeneous system in terms of a one-to-one mapping with the non-interacting and homogeneous one.

Cross-Chapters contents. The Chapters organization illustrated so far defines the choice we have made about the textbook content mainstream. Then, the central content of each Chapter is accompanied by side-dish contents, though equally important. So for example, the description of metals, semiconductors, insulators, superconductors and magnetically ordered samples is spread all over the textbook and in each Chapter, from the structure of the crystal to that of electronic, magnetic, and vibrational excitations, from the transport to the optical properties, to the treatment of interactions within Density Functional Theory. Similarly, each Chapter contains a specification of the given central content as applied to semiconductor nanostructures. Each Chapter also includes a discussion of the experimental methods suited to measure the given properties. A few exceptions are in order for space reasons: for example, the optical properties or the Density Functional treatment of magnetically ordered materials is not really discussed.

Cross-Chapter strategies. The following strategies are generally adopted across each Chapter.

(i) Ideas and concepts are developed from simpler to more complex situations. For example, the properties of infinite crystals are discussed following the evolution from the nanoscopic system of a few atoms in a chain, towards the full three-dimensional system. Or else, the phenomenological or macroscopic view of phenomena is discussed before entering any more formal development. This strategy is aimed to establish a tutorial approach.

(ii) Whenever possible the processes leading to the understanding of a given concept or phenomenon are explicitly discussed, so to highlight the difficulties encountered and how they have been solved. This is aimed to reinforce via real examples the idea that the working method in science naturally proceeds through cycles of theoretical hypothesis and experimental testing.

(iii) The manifestation of same concepts in different systems is underlined whenever possible, so to reinforce the ability of problem solving and of simplifying the amount of essential knowledge, via the setting of similar solutions to similar recognized problems.

(iv) Each Chapter tries to provide different and intersecting views of the given subject, in particular phenomenological, microscopic-theoretical, experimental, and ap-

plicative perspectives. This point-of-view mixing strategy is aimed to: reinforce the use of scientific method; make the discussion interesting and motivating at the same time to physicists, chemists, and engineers, each of them benefiting from the others perspectives; diversify the formal representations of the same subject and the quality of communication channels, so to lower the impedance in the process of knowledge sharing.

Cross-Chapters format. Coherently with the general and specific goals illustrated above and in order to provide tools oriented to empower the learning environment, the format of each Chapter is featured as follows.

A. The Chapter's introduction contains: in the form of questions, everyday life manifestations of the concepts discussed in the Chapter; a reasoned discussion about the links with previous and next chapters, in order to put everything in context.

B. Similarly, sections are devised with introductions explicitly discussing the logical and physical links with the adjacent ones. The A and B features are aimed to ease the student in gradually organizing the acquired knowledge in one picture.

C. Three types of content structures are introduced, that are immediately recognizable by different combinations of colors and shapes: Definitions and Properties, Concepts followed by Quick Questions, Procedures followed by Examples. Each type of content is logically specified by a rationale principle and by applications. So for example, for each Concept one or more conceptual Quick Questions are proposed to check the understanding, that is questions that can be answered with at most back-of-the-envelope calculations. For each type of problem to be solved a Procedure is coded and one or more Examples are worked out. The use of these structures is aimed to: focus the single significant concepts and tools composing the whole picture; favor the awareness that knowledge is acquired only once its conceptual, procedural, and factual components are metabolized, that is to say that no knowledge is possible without application from brain to hands and back.

D. The graphical appearance of figures, schematic pictures and sketches, plots of original experimental data, is especially taken care of and accompanied by self-contained captions. This is aimed to provide the student with an additional and diversified formal representation of the subject, other than wordy concepts or equations, for which understanding some students could be more talented than others.

E. Concepts and their connections, along with the procedures to solve typical problems are summarized at the end of each Chapter in a preliminary conceptual map. The student is then left with the task of realizing his/her own personalized conceptual map. This is aimed to ease the student in focusing the most significant issues and organizing her/his own knowledge.

F. Each last section is devoted to a description of experimental methods and strategies to probe the phenomena on which the Chapter is centered. This is aimed to provide the student with the essential tools and glossary to autonomously orient her/himself in the experimental literature. It is also aimed to help connecting calculated quantities with measured observables.

G. End-of-Chapter Problems with and without solutions are additional tools provided to further test the acquired procedures beyond the examples discussed in the main text, often staying within more elaborated situations.

H. End-of-Chapter Appendixes contain further details that would otherwise overburden the Chapter's development.

Acknowledgments

The authors represent three generations of students of a teacher, role-model in life and friend Franco Bassani, whom this textbook is dedicated to. In fact, the approach to a few selected subjects discussed in the present textbook is shared with his last textbook: F. Bassani and U.M. Grassano, *Fisica dello Stato Solido*, Bollati Boringhieri (Torino, 2000).

Special thanks are gratefully addressed to Prof. Domenico Ninno, who has inspired parts of this textbook and, in a preliminary draft, contributed with selected writing and figure setting.

Special thanks are also dedicated to Prof. Giuseppe La Rocca, for his critical reading of the manuscript and valuable suggestions.

The approach used to describe elasticity in solids and its connections with vibrational crystal modes is largely taken from the notes of a celebrated course that Fausto Fumi has given at the University of Genova and that one of us, G. I., knew.

Besides Bassani and Grassano's, other excellent textbooks have been quite inspiring in approaching selected subjects, and are indeed recurrently referred in the end-of-chapter bibliography. A special mention is devoted to the following textbooks oriented to Solid-State physics: Ashcroft and Mermin (*Solid-State Physics*, Holt-Saunders, Tokyo, 1976), Kittel (*Introduction to Solid State Physics*, John Wiley & Sons, Hoboken, 2005), Grosso and Pastori Parravicini (*Solid State Physics*, Academic Press, New York, 2000), Ziman (*Principles of the theory of solids*, Cambridge University Press, Cambridge, 1964). Selected subjects involving interacting electrons or superfluids are discussed also from the perspective of quantum liquids theories, where especially inspiring have been the books by Nozières and Pines (*The Theory of Quantum Liquids - Superfluid Bose Liquids*, Addison-Wesley, New York, 1990) and by March and Tosi (*Coulomb Liquids*, Academic Press, London, 1984). Baym's pedagogical lecture on superfluidity published in *Mathematical Methods in Solid State and Superfluid Theory* (Clark and Derrick eds., Oliver and Boyd, Edinburgh, 1969) is an essential easy-to-read reference for beginners and not. The description of binding mechanisms, transport and optical properties are analyzed in the significant books by Rohrer (*Structure and Bonding in Crystalline Materials*, Cambridge Univ. Press, Cambridge, 2004), Guozhong Cao (*Nanostructures and nanomaterials*, Imperial College Press, London, 2004), Datta (*Quantum transport*, Cambridge Univ. Press, 2005), Gaponenko (*Optical Properties of Semiconductors Nanocrystals*, Cambridge Univ. Press, Cambridge, 2005). For the connection with experiments, the pedagogical lecture by Martin (*Measurements and Correlation Functions in Many-Body Physics*, De Witt and Balian eds., Gordon and Breach, New York, 1968) represents a seminal reference, along with the classical textbook by Yu and Cardona (*Fundamentals of Semiconductors*, Springer, Berlin, 1996). Last but not least, the begin-of-chapter everyday-life phenomena and a number of re-

lated quick questions in the main Chapter's body, have found an inestimable inspiring source in the unique textbook by Bloomfield (*How Everything Works. Making physics out of ordinary*, John Wiley and Sons, USA, 2008), that we strongly suggest any student to go through.

We would also like to publicly appreciate the collaboration of ours Editor, who has closely followed us while defining the textbook format and has supported our proposed changes on the originally agreed terms, such as the introduction of the content structures and the use of a second color. We are convinced that these new ingredients have considerably contributed to improve the readability and possible success of this work.

Napoli and Pisa, March 2014

*Giuseppe Iadonisi
Giovanni Cantele
Maria Luisa Chiofalo*

Contents

| | | |
|----------|--|----------|
| 1 | Crystals and Nanosystems Structure | 1 |
| 1.1 | Introduction and basic definitions | 1 |
| 1.2 | Crystals and nanosystems structure | 6 |
| 1.2.1 | Bravais lattices representation | 6 |
| 1.2.2 | Lattice and crystalline symmetries | 9 |
| 1.2.3 | Nanocrystals | 22 |
| 1.3 | Lattice structure of specific crystals | 24 |
| 1.3.1 | Crystal structure of elements | 24 |
| 1.3.2 | Crystal structure of compounds | 29 |
| 1.4 | Reciprocal lattice | 32 |
| 1.4.1 | Definition of reciprocal lattice | 33 |
| 1.4.2 | Properties of the reciprocal lattice | 34 |
| 1.5 | Lattice planes | 38 |
| 1.5.1 | An useful theorem | 39 |
| 1.5.2 | Miller indexes | 40 |
| 1.6 | Experimental methods to determine crystal structures | 41 |
| 1.6.1 | Introduction | 41 |
| 1.6.2 | Scattering of radiation and particles from crystals: basic facts | 42 |
| 1.6.3 | Calculation of scattered intensities | 44 |
| 1.6.4 | Selected examples of scattering spectra | 46 |
| 1.6.5 | Bragg's law | 49 |
| 1.6.6 | Determination of the structure of crystals | 51 |
| 1.6.7 | Determination of the structure of nanosystems | 51 |
| 1.7 | Cohesion in solids | 57 |
| 1.7.1 | Phenomenology | 58 |
| 1.7.2 | Van der Waals, dipolar, and hydrogen bonds | 59 |
| 1.7.3 | Ionic bonds | 65 |
| 1.7.4 | Covalent bonds | 70 |
| 1.7.5 | Metallic bonds | 75 |
| 1.7.6 | Crystal bonding and periodic table | 78 |
| 1.8 | Mechanical properties of materials and crystal structures | 79 |

| | | |
|----------|---|-----------|
| 1.8.1 | Crystalline materials | 79 |
| 1.8.2 | Non-crystalline materials | 84 |
| | Summary of concepts and procedures | 86 |
| 1.9 | Appendix. Calculation of the transmission coefficient | 88 |
| 1.10 | Appendix. Energy and wavenumber of free particles | 89 |
| | Problems with solutions | 90 |
| | Problems without solutions | 95 |
| | References | 96 |
| 2 | Electronic structure of nanosystems and crystals | 97 |
| 2.1 | Introduction | 97 |
| 2.2 | Nanocrystals in one dimension | 99 |
| 2.2.1 | Tight-Binding method | 100 |
| 2.2.2 | Matrix elements | 102 |
| 2.2.3 | Electronic structure | 103 |
| 2.2.4 | Tight Binding with overlap | 109 |
| 2.2.5 | Surface and impurities states | 110 |
| 2.2.6 | Chain with $2N$ atoms of two different species | 116 |
| 2.2.7 | Applications to chains of metallic atoms | 117 |
| 2.2.8 | Evolving from nanocrystals to infinite crystals | 120 |
| 2.3 | Electronic structure of periodic crystals | 123 |
| 2.3.1 | Bloch theorem | 124 |
| 2.3.2 | Periodic boundary conditions | 126 |
| 2.3.3 | Forbidden and allowed bands | 133 |
| 2.3.4 | The effect of dimensionality through examples: the cases of graphene and carbon nanotubes | 140 |
| 2.4 | Methods for band-structure calculations | 146 |
| 2.4.1 | Plane Wave method | 146 |
| 2.4.2 | Wannier method | 154 |
| 2.4.3 | Pseudopotential methods | 156 |
| 2.5 | Many-particle effects: a first study | 161 |
| 2.5.1 | Interactions with ions: the Born-Oppenheimer approximation | 162 |
| 2.5.2 | Interactions among electrons: Hartree and Hartree-Fock methods | 164 |
| 2.5.3 | Electron affinity and ionization energies | 168 |
| 2.5.4 | Determination of electronic spectrum in the field of nuclei: jellium model | 171 |
| 2.5.5 | Determination of electronic bands of metals, semiconductors and insulators | 173 |
| 2.5.6 | Excitonic effects | 179 |
| 2.6 | Band structures in quantum confined nanostructures | 182 |
| 2.6.1 | Single-particle states | 182 |
| 2.6.2 | Excitons in nanometric systems | 187 |
| 2.7 | Spin-related magnetic ground states and excitations | 190 |
| 2.7.1 | Introduction and basic facts | 190 |

| | | |
|----------|--|------------|
| 2.7.2 | Diamagnetism | 193 |
| 2.7.3 | Paramagnetism | 195 |
| 2.7.4 | Ferromagnetism | 197 |
| 2.7.5 | Antiferromagnetism and other magnetic orderings | 201 |
| 2.7.6 | Structure of excitations in ferromagnets and antiferromagnets | 203 |
| 2.8 | Pair-correlated ground states and excitations | 207 |
| 2.8.1 | Superfluidity, superconductivity, Bose Einstein Condensation | 207 |
| 2.8.2 | Phenomenology of superconductors | 212 |
| 2.8.3 | Phenomenology of superfluids and the two-fluid model | 217 |
| 2.8.4 | BCS theory | 221 |
| 2.9 | Experimental methods to determine electronic spectra in crystals and nanostructures | 228 |
| 2.9.1 | Transport processes | 230 |
| 2.9.2 | Electron Energy Loss Spectroscopy | 230 |
| 2.9.3 | Photo-Emission Spectroscopy | 230 |
| 2.9.4 | Rayleigh Scattering Spectroscopy | 232 |
| 2.9.5 | X-ray Absorption Spectroscopy | 232 |
| 2.9.6 | Luminescence and X-ray Emission Spectroscopy | 233 |
| 2.9.7 | Auger Spectroscopy | 235 |
| 2.9.8 | Spectroscopies at larger energies | 236 |
| 2.9.9 | Electron Spin Resonance | 236 |
| 2.9.10 | Neutron scattering | 236 |
| | Summary of concepts and procedures | 237 |
| 2.10 | Appendix. Elements of variational calculus | 242 |
| 2.10.1 | Variational theorem | 244 |
| 2.11 | Appendix. Calculation of matrix elements within the Hartree-Fock approximation | 245 |
| 2.12 | Appendix. Electronic bands calculation in the jellium model | 247 |
| 2.13 | Appendix. Bose-Einstein condensation: basic concepts and more recent realizations | 248 |
| 2.14 | Representation of many particle fermion and boson states | 250 |
| 2.14.1 | Fermion states | 250 |
| 2.14.2 | Boson states | 251 |
| | Problems with solutions | 253 |
| | Problems without solutions | 262 |
| | References | 263 |
| 3 | Elements of continuum mechanics and lattice vibrations | 265 |
| 3.1 | Introduction | 265 |
| 3.2 | Macroscopic picture: deformations and stresses in continuous media | 267 |
| 3.2.1 | Rigidity, compressibility, and shape elasticity | 270 |
| 3.2.2 | Dynamics of continuous media and elastic wave propagation | 275 |
| 3.3 | Piezoelectricity and piezomagnetism | 283 |
| 3.3.1 | General facts | 284 |
| 3.3.2 | Applications | 285 |

| | | |
|--------|--|-----|
| 3.4 | Microscopic picture: normal modes | 286 |
| 3.5 | Vibrational modes of systems with finite number of atoms | 289 |
| 3.5.1 | Linear chain with N identical atoms interacting with first neighbors via elastic forces | 290 |
| 3.5.2 | Linear atomic chain with free edge atoms | 290 |
| 3.5.3 | Linear atomic chain with fixed edge atoms | 292 |
| 3.5.4 | Chain of $2N$ atoms with two alternating different species and first-neighbor interactions | 294 |
| 3.6 | Comments on systems with finite size | 297 |
| 3.7 | Vibrational modes of crystals | 298 |
| 3.7.1 | One dimensional crystal with one atom per cell | 299 |
| 3.7.2 | One-dimensional crystals with two atoms per cell | 303 |
| 3.7.3 | Three-dimensional lattices | 307 |
| 3.8 | Quantization of normal modes | 309 |
| 3.8.1 | Quantum nature of vibrational modes | 309 |
| 3.8.2 | Density of vibrational and electronic states | 312 |
| 3.9 | Thermodynamics of vibrational states: harmonic effects | 318 |
| 3.10 | Thermodynamics of vibrational states: anharmonic effects | 324 |
| 3.10.1 | Corrections to the specific heat in classical limit | 324 |
| 3.10.2 | Thermal expansion | 325 |
| 3.10.3 | Thermal conductivity | 326 |
| 3.11 | Electron-phonon interaction | 327 |
| 3.11.1 | General form of the electron-phonon interaction | 328 |
| 3.11.2 | Attractive interaction between electrons in metals | 330 |
| 3.12 | Experimental methods to determine vibrational spectra | 332 |
| 3.12.1 | Sound propagation | 333 |
| 3.12.2 | Inelastic Neutron Scattering | 334 |
| 3.12.3 | Infrared Spectroscopy | 337 |
| 3.12.4 | Raman Spectroscopy | 338 |
| | Summary of concepts and procedures | 342 |
| 3.13 | Appendix. Kinetic and potential energies of ions in harmonic approximation | 345 |
| 3.14 | Appendix. Calculation of normalization coefficients | 345 |
| 3.14.1 | Case of one atom per cell | 345 |
| 3.14.2 | Case of two atoms per cell | 346 |
| 3.15 | Appendix. Bloch theorem | 347 |
| 3.16 | Appendix. Calculation of free energy, entropy, and total vibrational energy | 348 |
| 3.17 | Appendix. Band calculation of the vibrational modes of crystals with one atom per cell | 349 |
| 3.18 | Appendix. Transition probabilities to first and second perturbative order | 351 |
| 3.19 | Appendix. Effective expression of the electron-phonon Hamiltonian H_{e-ph} | 357 |
| 3.19.1 | Calculation of the matrix element $F(\mathbf{q}, s, \mathbf{r})$ | 357 |

| | |
|---|------------|
| Problems with solutions | 358 |
| Problems without solutions | 368 |
| References | 368 |
| 4 Transport and Equilibrium Properties | 371 |
| 4.1 Introduction | 371 |
| 4.2 Equations of motion for electrons in a crystal and in the presence of external fields | 373 |
| 4.2.1 Motion in a static electric field | 376 |
| 4.2.2 Motion in a static magnetic field | 384 |
| 4.2.3 Motion in the presence of orthogonal electric and magnetic fields | 388 |
| 4.2.4 A generalized Newton law: the effective mass tensor | 389 |
| 4.2.5 Holes vs. electrons | 392 |
| 4.2.6 Introducing the scattering | 396 |
| 4.3 Metals | 401 |
| 4.3.1 Electric conductivity | 404 |
| 4.3.2 Thermal conductivity and thermoelectric effects | 406 |
| 4.3.3 Transport properties of metals within the Drude model | 411 |
| 4.3.4 Comparison with experiments requires Fermi-Dirac statistics: the Sommerfeld theory | 420 |
| 4.3.5 Drude model vs. Sommerfeld theory | 424 |
| 4.4 Semiconductors | 437 |
| 4.4.1 Chemical potential in intrinsic and doped semiconductors | 438 |
| 4.4.2 Intrinsic semiconductors | 443 |
| 4.4.3 Extrinsic semiconductors | 445 |
| 4.4.4 Donor and acceptor states | 446 |
| 4.5 Superconductors | 455 |
| 4.5.1 Conductivity and superconducting flow | 455 |
| 4.5.2 Specific heat | 458 |
| 4.5.3 Magnetic behavior | 458 |
| 4.6 Microscopic origin for the dependence of transport behavior on temperature | 458 |
| 4.6.1 Scattering with impurities or defects | 459 |
| 4.6.2 Scattering with phonons | 459 |
| 4.6.3 Scattering with other charge carriers | 460 |
| 4.6.4 Metals | 461 |
| 4.6.5 Semiconductors | 463 |
| 4.7 Devices based on $p-n$ junctions | 468 |
| 4.8 Electric conduction in nanostructures | 475 |
| 4.8.1 Phenomenology | 475 |
| 4.8.2 Conductance in nanostructures | 478 |
| 4.8.3 Current-drain and current-gate voltage characteristics | 482 |
| 4.9 Experimental methods to determine transport behaviors | 484 |
| Summary of concepts and procedures | 485 |

| | | |
|----------|---|------------|
| 4.10 | Appendix. Sommerfeld integrals | 488 |
| 4.11 | Appendix. Landau levels | 489 |
| 4.12 | Appendix. Fermi-Dirac integrals | 490 |
| 4.13 | Appendix. Fermi-Dirac statistics for impurity states in semiconductors | 492 |
| 4.14 | Appendix. Hall coefficient in semiconductors | 493 |
| 4.15 | Appendix. Currents from minority carriers in $p - n$ junctions | 494 |
| | Problems with solutions | 498 |
| | Problems without solutions | 505 |
| | References | 506 |
| 5 | Optical Properties | 509 |
| 5.1 | Introduction | 509 |
| 5.2 | Response to electromagnetic fields | 511 |
| 5.2.1 | A simple useful model | 511 |
| 5.2.2 | Dielectric function, polarizability and conductivity | 518 |
| 5.2.3 | Applications and comparison with experimental results | 527 |
| 5.3 | Quantum models for the dielectric function | 538 |
| 5.3.1 | Longitudinal dielectric function | 539 |
| 5.3.2 | Transverse dielectric function | 542 |
| 5.3.3 | Applications of the longitudinal dielectric function | 544 |
| 5.3.4 | Applications of the transverse dielectric function | 547 |
| 5.4 | Interaction effects | 549 |
| 5.4.1 | Screening of electron-electron interactions | 551 |
| 5.4.2 | Screening of ion-ion interactions | 555 |
| 5.4.3 | Effects of electron-phonon interaction on the optical properties | 557 |
| 5.5 | Optical properties of nanostructures | 565 |
| 5.5.1 | Quantum wells | 565 |
| 5.5.2 | Quantum dots | 568 |
| 5.6 | Experimental methods to determine optical properties | 569 |
| | Summary of concepts and procedures | 572 |
| 5.7 | Appendix. Definition of $\epsilon(\mathbf{k}, \omega)$ | 575 |
| 5.8 | Appendix. Analytic complex functions and differentiation | 579 |
| 5.8.1 | Integrals of complex-variable functions | 580 |
| 5.8.2 | Improper integrals | 581 |
| 5.9 | Appendix. Lindhard dielectric function | 583 |
| 5.10 | Appendix. Sum rules and dielectric functions | 583 |
| | Problems with solutions | 584 |
| | Problems without solutions | 590 |
| | References | 591 |
| 6 | Correlations and Density Functional Theory | 593 |
| 6.1 | Introduction | 593 |
| 6.2 | Exchange and correlation in N -particle systems | 594 |

| | | |
|--------|--|-----|
| 6.2.1 | Case with two particles | 595 |
| 6.2.2 | Case with N particles | 597 |
| 6.2.3 | The exchange and correlation hole | 599 |
| 6.3 | Self-consistent Hartree-Fock potential | 602 |
| 6.4 | Energy depends on density | 606 |
| 6.4.1 | Non interacting particles | 606 |
| 6.4.2 | The adimensional parameter r_s as a measure of density and coupling strength | 608 |
| 6.4.3 | Interacting particles: direct Hartree contribution | 610 |
| 6.4.4 | Interacting particles: exchange Fock contribution | 611 |
| 6.5 | Introducing spatial inhomogeneity via local density | 614 |
| 6.5.1 | Kinetic and interaction energy: Thomas-Fermi approximation | 614 |
| 6.5.2 | Exchange-correlation and screening effects | 615 |
| 6.5.3 | Thomas-Fermi linear screening | 617 |
| 6.5.4 | Exchange energy of the homogeneous electron gas: the Thomas-Fermi-Dirac model | 625 |
| 6.5.5 | Correlation energy | 626 |
| 6.5.6 | Density-gradient corrections | 626 |
| 6.6 | The Density Functional Theory | 627 |
| 6.6.1 | Variational formulation of the theory | 628 |
| 6.6.2 | Kohn-Sham equations | 630 |
| 6.6.3 | Meaning of the Kohn-Sham eigenvalues and eigenstates | 632 |
| 6.7 | Approximations for the exchange-correlation functional | 633 |
| 6.7.1 | Local-Density Approximation (LDA) | 633 |
| 6.7.2 | Connecting the exchange-correlation functional and the exchange and correlation hole | 635 |
| 6.7.3 | Selected exact results | 637 |
| 6.8 | Connection between DFT and response function theory | 638 |
| 6.9 | Comparison with other methods | 640 |
| 6.9.1 | Comparison with Hartree and Hartree-Fock models | 641 |
| 6.9.2 | Comparison with the Thomas-Fermi model | 642 |
| 6.10 | Applications of the DFT | 642 |
| 6.10.1 | Calculation of band structures: the cases of Si, GaAs and Graphene | 642 |
| 6.10.2 | Calculation of crystal dielectric functions | 643 |
| 6.10.3 | Application of DFT to electronic and optical properties in semiconductor nanostructures: the case of a quantum dot | 645 |
| 6.11 | Time-Dependent Density Functional Theory | 650 |
| 6.11.1 | Introduction: the Navier-Stokes equations | 651 |
| 6.11.2 | Time Dependent-DFT and microscopic Navier-Stokes equations | 654 |
| | Summary of concepts and procedures | 659 |
| 6.12 | Appendix. Virial theorem | 661 |

| | |
|---|------------|
| 6.13 Appendix. Screening effects in insulators within the Thomas-Fermi model | 662 |
| Problems with solutions | 664 |
| References | 671 |
| Subjects index | 673 |

Chapter 1

Crystals and Nanosystems Structure

Abstract The structure of crystals is here described after introducing Bravais and reciprocal lattices. Experimental techniques like X-ray elastic scattering and electron spectroscopy are briefly discussed, along with techniques specifically suited to nanostructures. The concept of cohesive energy is introduced and a few related bond mechanisms are reviewed: Van der Waals, ionic, covalent, and metallic.

1.1 Introduction and basic definitions

Did you ever ask yourself why diamond is so precious, and graphite yet composed of Carbon as well, is not. Why the former is hardest and the latter exfoliates. Why knives are eventually made with steel, though cooks could sometimes prefer iron-made ones. Why glasses or plastics have so different behavior under strains or stresses [1]. How do we picture the general idea that crystals are made of atoms arranged in regular lattices, and how shall we choose the kind of experiment to probe these crystal structures. How those beautiful pictures of single atoms are eventually taken, if the wavelength of visible light, orders of microns, is thousands times longer than an atom size, in the Å range. What sticks a crystal together and which is the connection with ordinary chemistry and nature of chemical bonds. And, after all, is there a structured theoretical knowledge where all the answers to these questions are hardwired together, along with experimental concepts and techniques scientifically grounding it: crystal structure, nature of chemical bonds, mechanical properties and on-purpose engineering of materials. This Chapter is conceived to open up useful answers.

The description of solid structures is a fundamental part in any textbook of solid-state physics, since several physical properties of materials are actually traced back to it. Textbooks sometimes devote even more than one Chapter to this topic, essentially sharing a common layout [2, 3, 4]. It is well known that a solid might aggregate under different structures: crystalline, polycrystalline, amorphous. Each of them is

unambiguously determined by a set of distinguishing parameters. Here follow the basic definitions.

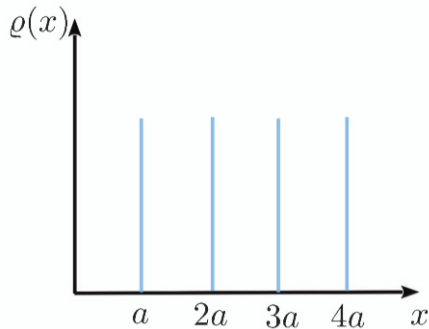
Definitions of crystalline, polycrystalline and amorphous structures. In solid-state physics the name crystal is usually referred to crystalline structures.

Definition

Crystalline structures are characterized by the regular arrangement in space of single atoms or collections of them, leading to long-range order.

Periodicity in matter distribution introduces significant consequences. For example, electrons and ion vibrations can be represented as stationary waves, therefore characterized by discrete allowed frequencies: much alike a violin string that - when struck - can vibrate at the fundamental tone frequency and overtones that are compatible with the specific string length, thickness, material composition. These allowed energies can in fact be represented as bands that are periodic in wavenumber and that take into account, to some approximation, many-body interactions: electrons and ion vibrations can indeed be viewed as if they were single particles in quantum states characterized by these energy bands, an example of the seminal concept of collective excitation that is deepened later on in this textbook.

Fig. 1.1 Representation of crystal ordering for a one-dimensional chain of atoms evenly spaced by a . The typical distribution function $\rho(x)$ is sketched as a function of the distance x from a reference atom located at the origin



Definition

Polycrystalline structures are composed of an ensemble of crystalline structures extending over a finite spatial range, with adjacent structures being oriented in different ways.

The physics describing this situation depends on the single-crystals sizes. It includes the two limiting cases of an infinite crystal and of a nanostructure, while the size of the single crystals become sufficiently large or small, respectively. Finally,

Definition

Amorphous structures are characterized by a spatial ordering limited to the vicinity of the single atom or molecule.

Glass materials are one example of amorphous structures. They are usually produced when a viscous molten material is so rapidly cooled below a given (glass) transition temperature, that atoms have not enough time to arrange in a seed crystal structure and then grow up from it: viscosity can be as large as 10^{12} Pa·s, so that atoms jerk around but with no possibility of moving one past the other. Plastics materials like polyethylene are one more example, with one important difference: they are based on polymers, that are long chains of thousands atoms or molecules, and these are arranged partly in crystal - say, 80% - and partly in amorphous structures.

Concept

Analytical and graphical representations of this concept of ordering are provided by the so-called distribution function $\rho(\mathbf{r})$, yielding the number of atoms per unit volume as a function of the distance r from the given reference atom.

In one dimension (1D) the distribution function is

$$\rho(x) = \sum_{n=-\infty}^{\infty} \delta(x - na) ,$$

that is a sum of $\delta(x)$ Dirac functions centered on lattice sites. This is sketched in Fig. 1.1: in 1D the number of first, second, and so on neighbors is the same, and so is the peaks height. This is not the case in three dimensions: Fig. 1.2 depicts a typical $\rho(r)$ for the case when atoms are arranged and located at the vertex of adjacent cubes infinitely replicated to fill in the space. In fact, the number of neighbors in a spherical shell of radius Na and width dr is

$$dN = \frac{4\pi(Na)^2}{a^3} dr .$$

Thus, the number of neighbors of order N per unit length is

$$\frac{dN}{dr} = \frac{4\pi N^2}{a} . \quad (1.1)$$

In a two-dimensional square lattice, the analogue of (1.1) is

$$\frac{dN}{dr} = \frac{2\pi N}{a} . \quad (1.2)$$

Note now that while (1.1) scales with N^2 , (1.2) scales linearly with N .

Notice that the number of first neighbors of order N increases indefinitely, so that the probability of finding one particle at large distance from a given reference

particle tends to be 1. The probability of finding a particle at distance r from another selected particle located at the origin $r = 0$ is represented by the so-called radial pair correlation function $g(r)$: this is a significant concept indeed, since the probability depends on the correlations between particles. In essence, in a uniform system with density n , $ng(r)$ can be viewed as the local density which could be observed if one chose to sit on top of a particle at the origin, or else $ng(r)4\pi r^2 dr$ the number of particles in a shell dr thick and placed at distance r from the origin. A typical behavior of the radial pair correlation function is sketched in Fig. 1.3 for either liquid or amorphous structures.

Infinite crystals, single crystals and nanostructures. Whether a material solidifies into either a crystalline or polycrystalline, or else amorphous structure, is a complicated issue strongly affected by conditions of pressure, temperature, impurity content, and so on. While polycrystalline or amorphous materials are most often the result of spontaneous growing, the growth of a single-crystal is almost always a very laborious task, requiring efficient and accurate control of the above parameters. Progress along these lines has led to the revolutionary ability of engineering the so-called nanostructures.

Definition

A crystalline nanostructure is one single tiny crystal whose size may vary from about ten to one hundred Angstrom: namely from the size of one single molecule to the typical size of a structure where long-range crystal ordering begins to show up.

The physics describing these structures is quite elaborate, since it involves effects due to finite size. It also depends on the type of material, besides the environmental parameters. Size-dependent properties show up in nanostructures, which are absent in macroscopic infinite crystals. As discussed in Chap. 2, the bulk properties of a macroscopic crystal can be described in terms of those for an infinite crystal, rather than in terms of a large but finite one, provided that the total number of atoms be much larger than the number of surface atoms. Consider for example the cubic lattice with spacing a , which Fig. 1.2 refers to. Be $L = 1000a$ the size

Fig. 1.2 Representation of crystal ordering for a three-dimensional crystal made of atoms located at the vertex of adjacent cubes with corner a indefinitely replicated to fill in the space. The typical distribution function $\rho(r)$ is sketched vs. the distance r from a reference atom located at the origin

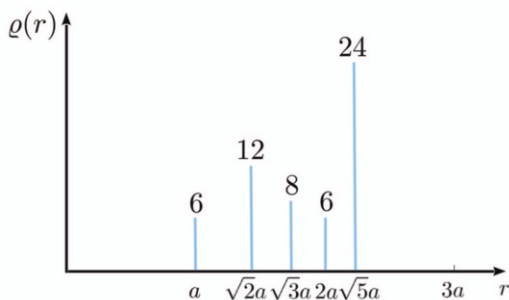
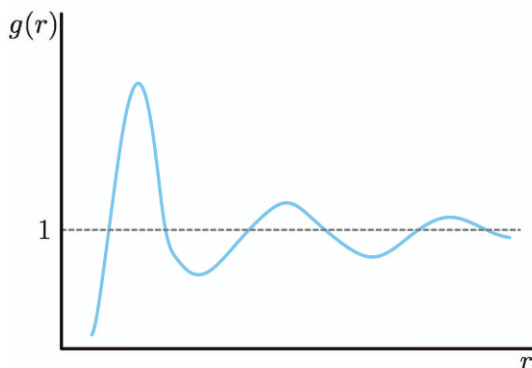


Fig. 1.3 Representation of crystal ordering for a three-dimensional liquid or amorphous structure. The typical correlation function $g(r)$ for liquid or amorphous structures is sketched vs. the distance r from a reference atom located at the origin



of the macroscopic sample and $N = 1000^3$ the total number of atoms, so that the surface atoms are approximately in the number of $N_S \simeq 6 \cdot 1000^2$. One therefore has $N_S/N \simeq 6/1000$, that is the surface atoms are about 0.6% of the total number of atoms and the limit of infinite crystal can be considered valid.

Chapter's organization. The Chapter is organized as follows. First, concepts and theoretical tools are given, that are useful to describe those characteristics of solids occurring in crystalline structures as well as small-size systems. Then, the comprehension of this involved matter is expanded while discussing the basic experimental methods that are used to determine those structures, such as X-ray scattering and electron spectroscopy. While the understanding of the mechanisms leading to a specific crystal formation is founded on a few bold principles, the development of a general theory able to explain and predict detailed behaviors is a quite challenging task. In fact, given an atom aggregate within a finite volume, it is well known that the conditions which realize thermodynamical equilibrium are determined after minimizing the total free energy at finite temperature T , or the total energy at $T = 0$. Therefore, the basic issue here is the calculation of the total energy. This can be accomplished for specific systems only by means of powerful computational tools. A central concept in this game is the cohesive energy, that is the energy required to take one atom or molecule away from the solid, towards infinity. The final part of the Chapter is devoted to discuss the determination of cohesive energy for different materials within tractable approximations, and to trace back the phenomenology of chemical and mechanical properties.

1.2 Crystals and nanosystems structure

1.2.1 Bravais lattices representation

On a macroscopic scale, a crystal can be defined as a solid limited by planar faces at well-specified angles, whose chemical composition results to be uniform. Our present knowledge of crystal structure has been built up at the beginning of the 20th Century after the discovery of X-rays diffraction phenomena in the very first experiments of von Laue. Earlier studies of crystals were based on simple observations about regular appearance and outer shapes of materials.

Up-to-date techniques have allowed to set in a correct definition of crystal structure. In particular, a crystal is composed after infinite periodic repetitions in space of replicas of one and the same structural unit. An unambiguous definition of crystal requires the setting of two fundamental concepts, as sketched in the left panel of Fig. 1.4:

Definition

Bravais lattice. It is the infinite set of geometrical points, arranged in space according to a regular and periodic manner. These points are named lattice sites. They share the following property: the geometrical description of the material, including position, orientation and type of atoms, looks always the same from the perspective of whatever site one might choose as point of sight.

Base. It is the basic structure unit, composed of one or more atoms, molecules and/or ions. Its chemical composition may run from one single atom in crystals of gold and alkaline metals to tens, hundreds or thousands atoms in inorganic and organic crystals, and up to a hundred-thousand atoms in protein-based crystals such as DNA.

A real crystal is obtained after associating one base to each Bravais lattice site. Thus one has first to construct the Bravais lattice, a task which can be accomplished in three steps as in the following procedure:

Procedure

Step 1. Let \mathbf{O} be an arbitrary site chosen as reference or origin. Let a second site be chosen among the first neighbors and \mathbf{a}_1 be the vector identifying this second site from \mathbf{O} . The definition of lattice states that there are other lattice points located on the line along the vector \mathbf{a}_1 . The points on this line are represented by the vector $\mathbf{R} = n_1 \mathbf{a}_1$ with n_1 an integer number positive, negative, or zero.

Step 2. Consider now the first neighbors to \mathbf{O} , which do not belong to the straight

line in Step 1. Let us consider one among this second set of neighbors, and let \mathbf{a}_2 be the vector locating the corresponding site from the reference \mathbf{O} . The definition of lattice states that all the points in the plane represented by the vector $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2$, with n_1 and n_2 integer relative numbers, are in turn lattice points.

Step 3. Consider eventually the first neighbors to \mathbf{O} , which do not belong to the plane in Step 2. Let us consider one among this third set of neighbors, and let \mathbf{a}_3 be the vector locating the corresponding site from the reference \mathbf{O} . Applying one more time the definition of lattice, one may conclude that all of the sites in the three-dimensional (3D) lattice are represented by the vector

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \quad (1.3)$$

with n_1 , n_2 and n_3 relative integer numbers.

Concept

The Bravais lattices share the following fundamental idea: a rigid translation of the lattice by the vector \mathbf{R} given by Eq. (1.3) leaves the lattice unchanged.

That is, Bravais lattices are invariant with respect to rigid translations of a lattice vector. Equation (1.3) can be viewed either as the law which locates the lattice positions or as a law of translational invariance: \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 are called primitive translational vectors. Their relative direction, orientation, and size determine the specific type of Bravais lattice.

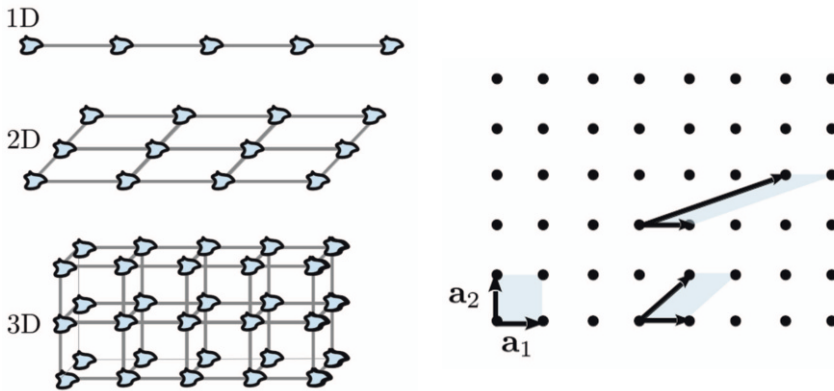


Fig. 1.4 Schematic representation of Bravais lattices with base. Left panel, from top to bottom: scheme for crystals in one (1D), two (2D), and three (3D) dimensions. The basic structure unit (base) is repeated at each site of the Bravais lattice. Right panel: non-univocal choice of the primitive cell. Three different possible choices of primitive cell for a 2D square lattice are sketched. Bottom left: a square. Bottom right: a rhomb. Top: a second rhomb with different side sizes

A further useful tool is obtained by the following construction. Let us trace from the base and the tip of each \mathbf{a}_i , with $i = 1, 2$ and 3 , the plane parallel to that determined by the remaining two \mathbf{a}_j ($j \neq i$). The resulting solid is called primitive cell. Primitive cells share two basic properties:

Properties

P1. The volume of the primitive cell is

$$V_c = |(\mathbf{a}_1 \wedge \mathbf{a}_2) \cdot \mathbf{a}_3| . \quad (1.4)$$

In the simple cases sketched in the left panel of Fig. 1.4, the primitive cells are represented by a segment in 1D, a square in 2D and a cube in 3D.

P2. When replicated next to one another, primitive cells fill in the whole space, leaving no room for either void space or superpositions.

Therefore:

Concept

A real crystal can be identified by providing the primitive cell and the base, namely the atomic species and the arrangement of different atoms within the primitive cell.

It is easily seen how a one-to-one correspondence exists between lattice sites and primitive cells in any given Bravais lattice. This is the case once a choice has been taken for the primitive cell. Indeed, the latter is by no means univocal: infinite equivalent selections can be picked up, all of them having the same volume size and properties of space filling.

The right panel of Fig. 1.4 depicts three different possible choices of primitive cell for a 2D square lattice. Simple geometrical considerations allow to conclude that all the three of them have the same area size.

Among all possible primitive cells, a special choice is named after Wigner and Seitz, and is constructed according to the following procedure:

Procedure

Step 1. Let \mathbf{O} be an arbitrarily chosen lattice point, serving as origin.

Step 2. Consider the vectors \mathbf{R} in (1.3), determining all the lattice points from the origin \mathbf{O} . For each of them trace the plane perpendicular to it and cutting the vector in halves.

Step 3. Trace the region of space around the reference lattice point \mathbf{O} , and bounded by the closest planes as determined in Step 2 above.

Definition

The primitive cell determined along the Steps 1–3 above is the elementary Wigner-Seitz cell.

Fig. 1.5 depicts steps 1–3 above for the case of the 2D square lattice of the right panel of Fig. 1.4 and for the 2D oblique lattice.

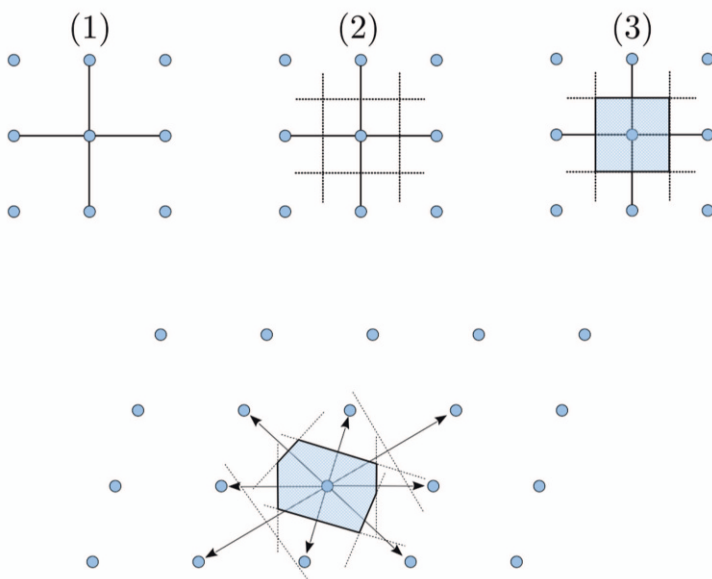


Fig. 1.5 Construction of the Wigner-Seitz cell. Top: the case of the 2D square lattice in the right panel of Fig. 1.4 is depicted. From left to right: the three steps described in the text. Bottom: the more general case of an oblique 2D lattice

The special significance of the Wigner-Seitz cell is matter of the following section.

1.2.2 Lattice and crystalline symmetries

The physical properties of materials are intimately related to the symmetry degree of the crystal structure. For example, crystals of Si have a higher degree of spatial

symmetry in pure form rather than in the quartz form, SiO_2 . In fact, the properties of the former are quite different from those of quartz: this is a piezoelectric material used in clocks, radios and digital devices, namely capable of generating an electric polarization from mechanical deformations and viceversa. Therefore, the investigation of crystal symmetries is a crucial issue in solid-state physics.

The Wigner-Seitz cell is a central concept here, since it can be shown that any lattice symmetry operation keeping a lattice point fixed, also leaves the whole Wigner-Seitz cell unchanged.

Definition

A coordinate transformation is said to be rigid if it leaves unaltered the distances between all Bravais lattice points

Definition

A symmetry operation of the Bravais lattice is a rigid coordinate transformation which leaves the Bravais lattice unaltered

Definition

A symmetry operation is said to be punctual if it leaves fixed at least one Bravais lattice site

Definition

The set of symmetry operations which transform the lattice in itself while keeping a fixed lattice point is called lattice point group.

Concept

The lattice-point group plays a very significant role, since any symmetry operation U can be obtained as the combination of a lattice-point group operation S followed by the rigid translation as in (1.3).

This is easily demonstrated as follows. Consider a lattice point \mathbf{O} as the origin. The operation U transforms the origin in \mathbf{R}_1 . The operation $T_{-\mathbf{R}_1}U$, that is U followed by the translation by $-\mathbf{R}_1$, leaves \mathbf{O} unchanged. Thus, $T_{-\mathbf{R}_1}U$ is by definition an operation of the lattice point group. Therefore one has $T_{-\mathbf{R}_1}U = S$ or else $U = T_{\mathbf{R}_1}S$, which demonstrates the statement. Here are the typical point-group symmetry operations:

Reflection with respect to a plane: each site is mirrored with respect to the given plane.

Inversion with respect to a point O: a point P identified by vector \mathbf{r} with respect to the inversion point O, is transformed into P' identified by $-\mathbf{r}$.

Rotation around a given axis.

The existence of translational symmetry fixes well-given requirements on the possible rotational symmetry operations. In fact, only rotations of angle $\varphi = 2\pi/n$ with $n = 1$ (around a so-called unary axis), 2 (binary), 3 (ternary), 4 (quaternary), and 6 (senary) are allowed. This is easily demonstrated after inspection of Fig. 1.6. Let A and B be two first neighbors lattice points and $a = AB$. Consider a rotation of angle φ around an axis perpendicular to a plane containing A and B: this is a symmetry of the crystal. A rotation of angle φ , around the rotation axis passing through A, transforms B into B'. In a similar operation, a rotation of angle $-\varphi$ around the rotation axis passing through B transforms A into A'. A' and B' are lattice sites, and thus the distance between them has to be an integer v multiple of the lattice spacing a . By inspection, one has:

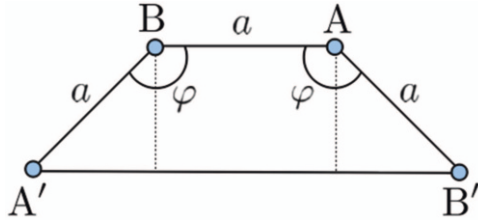
$$a + 2a \sin\left(\varphi - \frac{\pi}{2}\right) = a - 2a \cos(\varphi) = va,$$

that is

$$\cos(\varphi) = \frac{1 - v}{2}. \quad (1.5)$$

Since $|\cos(\varphi)| \leq 1$, (1.5) is valid only whenever $v = -1, 0, 1, 2$, and 3, corresponding to the values $\varphi = 2\pi/n$ with $n = 1, 2, 3, 4$, and 6.

Fig. 1.6 Translational symmetry restricts the number of possible rotations to angles $\varphi = 2\pi/n$, $n \leq 6$ excluding the case with $n = 5$. The geometric construction to find possible rotations is depicted



Concept

Lattice-symmetry operations limit the kind of allowed pavings. For example, translational and rotational symmetry operations in a lattice imply that it is impossible to pave a room floor with pentagonal tiles.

Definition

The set of symmetry operations belonging to the lattice point group together with the translational symmetries defines the space group of the Bravais lattice.

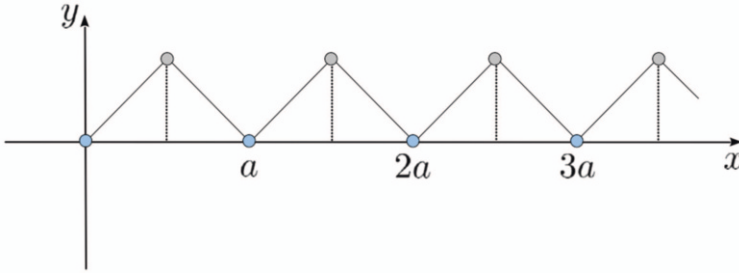


Fig. 1.7 Space-group operations in a Bravais lattice with base, showing that translations might be included, that do not correspond to a lattice vector. Sketched is the case of a one-dimensional lattice with two atoms per unit cell. The space-group operation $\{\alpha, \tau\}$ is composed of the point-group rotation α of π around the x -axis, followed by an oblique translation amounting to $\tau = (a/2, a/2)$

For a real crystal, the following notation is from now on used: $\{\alpha, \tau\}$ indicates a space-group operation composed of a point-group operation $\{\alpha\}$ and a translation $\{\tau\}$. A special operation is the identity E , thus $\{E, \mathbf{R}\}$ is just a translation and $\{\alpha, 0\}$ just a point-group operation. The point group might occur to contain one single operation, that is the identity.

The introduction of a base does not necessarily preserve the point-group symmetry operations and it has in general relevant consequences on the allowed crystal symmetries. Therefore possible crystalline space groups are only those which can be constructed, as in the previous case without base, after combining symmetry operations belonging to the point-group with translations. However, possible translations do not necessarily correspond now to a lattice vector $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$. Explanatory examples of this illuminating quibble are illustrated in Figs. 1.7 and 1.8. Here, cases are displayed of space-group operations containing either rotations or reflections, followed by translations that do not correspond to a lattice vector.

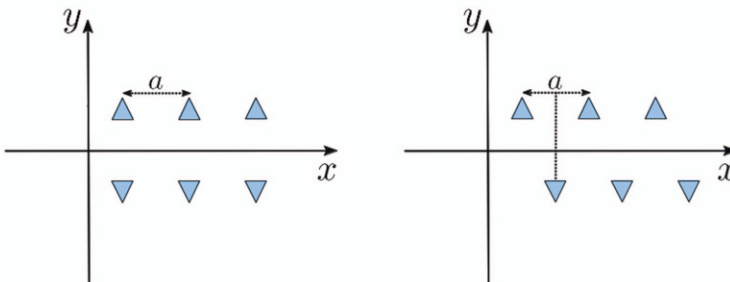


Fig. 1.8 Space-group operations in a Bravais lattice with base, showing that translations might be included, that do not correspond to a lattice vector. Sketched is the case of a 1D lattice with base. The space-group operation $\{\alpha, \tau\}$ is composed of a specular reflection α with respect to the x -axis in the left crystal, and in the right crystal by a specular reflection α with respect to the x -axis, followed by a translation by $\tau = (a/2, 0)$ along x

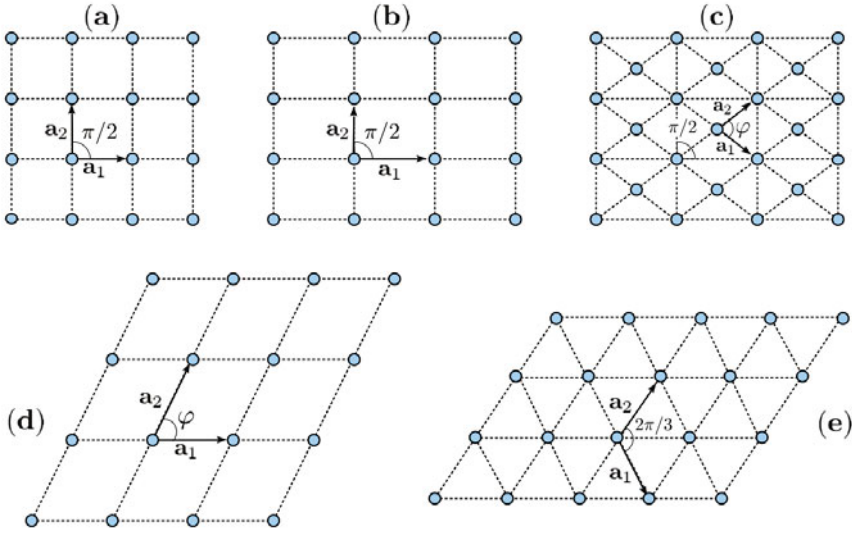


Fig. 1.9 Bravais lattice in 2D. The five conceivable Bravais lattices in 2D are depicted from (a) to (e) according to the description in the main text

According to the discussion above, it is now possible to list all the possible Bravais lattices. In the following a systematic illustration is given of the 5 Bravais lattices that are possible in 2D and the 14 possible in 3D, using the Procedures coded so far.

1.2.2.1 Bravais lattices in two dimensions

The five Bravais lattices that can be conceived in two dimensions are depicted in Fig. 1.9 along with the vectors defining the primitive cell.

Here they are described in terms of the elementary translation vectors \mathbf{a}_1 and \mathbf{a}_2 and of the angle φ that they form:

Properties

P1. Squared: $|\mathbf{a}_1| = |\mathbf{a}_2|$ and $\mathbf{a}_1, \mathbf{a}_2$ placed at angle $\varphi = \frac{\pi}{2}$.

P2. Rectangular: $|\mathbf{a}_1| \neq |\mathbf{a}_2|$, and $\varphi = \frac{\pi}{2}$.

P3. Body-centered rectangular: As in the rectangular case in P2, but an extra lattice point is located at the center of each rectangle.

Note that this is a cell with two lattice points, located at center and vertex positions of the rectangle. A cell including more than one lattice site and designed to have all the symmetries of the given lattice is usually named conventional cell. In particular, here the conventional cell is actually not the primitive cell. The latter can instead be represented considering the cell origin at the rectangle center and choosing \mathbf{a}_1 and \mathbf{a}_2 to be the vectors connecting the center with two vertexes of the same rectangle. In this representation, $|\mathbf{a}_1| = |\mathbf{a}_2|$ but $\varphi \neq \pi/2$.

P4. Slanted or oblique: $|\mathbf{a}_1| \neq |\mathbf{a}_2|$ and $\varphi \neq \pi/2$.

P5. Hexagonal: $|\mathbf{a}_1| = |\mathbf{a}_2|$ and $\varphi = 2\pi/3$.

The point-group symmetries related to these 5 lattices are easily picked out. Starting from that with lower symmetry, the slanted lattice in P4 has binary-axes symmetries. To the above binary-axes symmetries, the rectangular P2 and the body-centered rectangular P3 add reflections with respect to the straight-dashed lines in Fig. 1.9 (b) or any other straight line parallel to the latter and crossing at the rectangles centers. The squared lattice in P1 adds quaternary axes symmetries. Finally the hexagonal P5 has ternary and senary-axes symmetries, besides reflection symmetries. Note the peculiarity here that two different Bravais lattices, the rectangular and the body-centered rectangular, share the same point-group symmetry.

Quick Questions

Q1. List all the point-group symmetry operations for a squared lattice.

Why a centered squared lattice has not been listed?

Answer. Point-group symmetries are: rotations by $2\pi, 3\pi/2, \pi, \pi/2$; reflections: with respect to any axis parallel to x and y , passing through lattice points or the centers of the squares. The centered squared lattice has not been listed since it is equivalent to a squared lattice with side $a\sqrt{2}/2$, where a is the side of the original squared lattice.

Q2. List all of the symmetry point-group operations of the hexagonal lattice.

Answer. Hint: repeat the reasoning in the answer to Q1.

Q3. Say whether a hexagonal lattice with no points at the hexagon center can be a Bravais lattice.

Answer. No, since the lattice points cannot be equivalent to each other.

Indeed, consideration of two nearest-neighbor atoms is enough to understand that arrangement and orientation of the atoms do not appear equivalent when viewed from these two atoms perspective. Such a lattice can be described in terms of a hexagonal Bravais lattice with a two-atom basis.

1.2.2.2 Bravais lattices in three dimensions

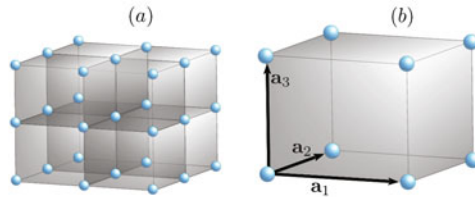
The 14 Bravais lattices that can be conceived in three dimensions can be classified into seven subgroups, according to the type of point-group symmetry. These are: cubic, tetragonal, orthorhombic, monoclinic, triclinic, trigonal, and hexagonal systems and are systematically illustrated in the following, along with the defining point-group operations and primitive vectors \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 . In classifying them, the introduction of the conventional cell results to be quite convenient, in a manner similar to the 2D rectangular lattices case.

Cubic system

The point-group symmetries are those which leave a cube unchanged and can be counted in the number of 48. Three different Bravais lattices are classified in the cubic system: simple cubic, body-centered cubic, and face-centered cubic. All of them are described below, since a great variety of materials occur in either one of these structures.

Simple cubic. It is amenable to the simplest and most intuitive description, as depicted in Fig. 1.10. The primitive cell in Fig. 1.10 (b) is indeed a cube with side a .

Fig. 1.10 Bravais lattices in 3D. Cubic system. The simple cubic lattice is depicted. (a) Finite portion of the lattice. (b) Primitive cell



Let the reference system be chosen with cartesian axes along the three orthogonal cube corners and represented by unitary vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$. One has:

$$\mathbf{a}_1 = a\hat{\mathbf{x}} \quad ; \quad \mathbf{a}_2 = a\hat{\mathbf{y}} \quad ; \quad \mathbf{a}_3 = a\hat{\mathbf{z}}. \quad (1.6)$$

The volume of the primitive cell is $V_c = a^3$. The number of first neighbors is six, located at $(\pm a, 0, 0)$, $(0, \pm a, 0)$, and $(0, 0, \pm a)$ with respect to the origin. The number of second neighbors is 12, located at $(\pm a, \pm a, 0)$, $(0, \pm a, \pm a)$, and $(\pm a, 0, \pm a)$. The third neighbors are 8 at positions $(\pm a, \pm a, \pm a)$.

Body-centered cubic. It is obtained from the simple cubic after inserting one lattice site at the center of the cube.

Fig. 1.11 Bravais lattices in 3D. Cubic system. The body-centered cubic lattice is depicted. (a) Conventional cell. (b) A possible choice of vectors \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 in the primitive cell

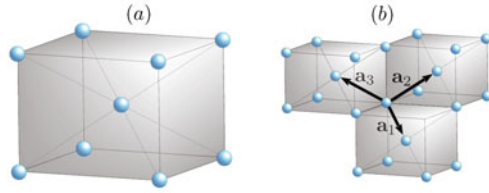


Fig. 1.11 displays in panel (a) the conventional cell, which includes two sites, and a possible choice of primitive vectors in panel (b). These are:

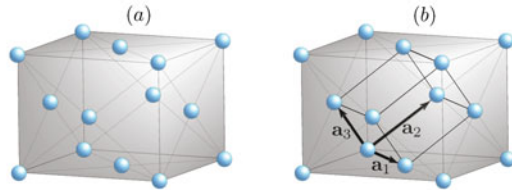
$$\mathbf{a}_1 = \frac{a}{2}(\hat{\mathbf{x}} - \hat{\mathbf{y}} - \hat{\mathbf{z}}) \quad ; \quad \mathbf{a}_2 = \frac{a}{2}(\hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}}) \quad ; \quad \mathbf{a}_3 = \frac{a}{2}(-\hat{\mathbf{x}} - \hat{\mathbf{y}} + \hat{\mathbf{z}}). \quad (1.7)$$

The primitive cell generated by the vectors (1.7) is a rhomboedron with corner size $\sqrt{3}a/2$ and angle $109^\circ, 28'$ between adjacent corners. Application of (1.4) with the use of (1.7) yields the volume size of the cell to be $V_c = a^3/2$. From this result, it is evident that the conventional cell with volume a^3 , indeed contains two elementary cells. The lattice has 8 first neighbors located at $(\pm a/2, \pm a/2, \pm a/2)$, 6 second neighbors at $(\pm a, 0, 0)$, $(0, \pm a, 0)$, and $(0, 0, \pm a)$, and 12 third neighbors at $(\pm a, \pm a, 0)$, $(\pm a, 0, \pm a)$, and $(0, \pm a, \pm a)$.

Note that the site at the cube center is all but special. Indeed, if the full lattice were drawn, one would see that the lattice sites at the cube center could be as well chosen as vertexes of a new conventional cell, with the earlier sites at the vertex being positioned at the center of the new cell. This is a manifestation of the Bravais-lattice concept, stating that a Bravais lattice looks the same from whatever lattice site is chosen as point of sight.

Face-centered cubic. It is obtained from the simple cubic lattice after inserting one lattice site at the center of each of the six faces.

Fig. 1.12 Bravais lattices in 3D. Cubic system. The face-centered cubic lattice is depicted. (a) Conventional cell. (b) A possible choice of vectors \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 in the primitive cell



The panel (a) of Fig. 1.12 displays the conventional cell. This now includes four sites, since the vertex points are shared among 8 cells and the face points are shared among 2 cells. A possible choice of primitive vectors is displayed in panel (b). These

are:

$$\mathbf{a}_1 = \frac{a}{2}(-\hat{\mathbf{x}} + \hat{\mathbf{y}}) \quad ; \quad \mathbf{a}_2 = \frac{a}{2}(\hat{\mathbf{y}} + \hat{\mathbf{z}}) \quad ; \quad \mathbf{a}_3 = \frac{a}{2}(\hat{\mathbf{z}} - \hat{\mathbf{x}}). \quad (1.8)$$

The primitive cell in Fig. 1.12 (b) is a rhomboedron. The volume size is $V_c = a^3/4$, showing in fact that the conventional cell corresponds to four primitive cells. The lattice has 12 first neighbors at $(\pm a/2, \pm a/2, 0)$, $(\pm a/2, 0, \pm a/2)$, and $(0, \pm a/2, \pm a/2)$, 6 second neighbors at $(\pm a, 0, 0)$, $(0, \pm a, 0)$, and $(0, 0, \pm a)$, and 24 third neighbors at $(\pm a, \pm a/2, \pm a/2)$, $(\pm a/2, \pm a, \pm a/2)$, and $(\pm a/2, \pm a/2, \pm a)$.

In conclusion, symmetry properties in the Wigner-Seitz cells of cubic lattices are not immediately at sight. This is the reason why the body- and face-centered cubic lattices are described by means of the conventional cell with volume a^3 and with two and four lattice sites, located at $(0,0,0)$ and $(a/2, a/2, a/2)$, and at $(0,0,0)$, $(a/2, a/2, 0)$ and $(a/2, 0, a/2)$, $(0, a/2, a/2)$, respectively.

Tetragonal system

A tetragonal lattice is originated by a cube that is transformed into a parallelepiped with square base. Two different Bravais lattices are classified in the tetragonal system: simple and body-centered tetragonal lattice.

Simple tetragonal. It is originated from the simple cubic lattice. The primitive vectors are:

$$\mathbf{a}_1 = a\hat{\mathbf{x}} \quad ; \quad \mathbf{a}_2 = a\hat{\mathbf{y}} \quad ; \quad \mathbf{a}_3 = c\hat{\mathbf{z}}. \quad (1.9)$$

The cell volume amounts to $V_c = a^2c$. The number of first, second, and third neighbors depends on the size values of the elementary-cell corners.

Body-centered tetragonal. It is originated from the body-centered cubic lattice. The base vectors are:

$$\mathbf{a}_1 = a\hat{\mathbf{x}} \quad ; \quad \mathbf{a}_2 = a\hat{\mathbf{y}} \quad ; \quad \mathbf{a}_3 = \frac{a}{2}(\hat{\mathbf{x}} + \hat{\mathbf{y}}) + \frac{c}{2}\hat{\mathbf{z}}. \quad (1.10)$$

The cell volume amounts to $V_c = a^2c/2$. The simple and body-centered tetragonal lattices are depicted in Fig. 1.13 (a) and (b), respectively.

Orthorhombic system

The orthorhombic system is originated from the tetragonal one, after relaxing the requirement that the base be squared. Four different Bravais lattices are classified in the orthorhombic system: simple, base-centered, body-centered, and face-centered orthorhombic lattice.

Simple orthorhombic. It is originated from the simple tetragonal as in Fig. 1.14 (a). The primitive vectors are:

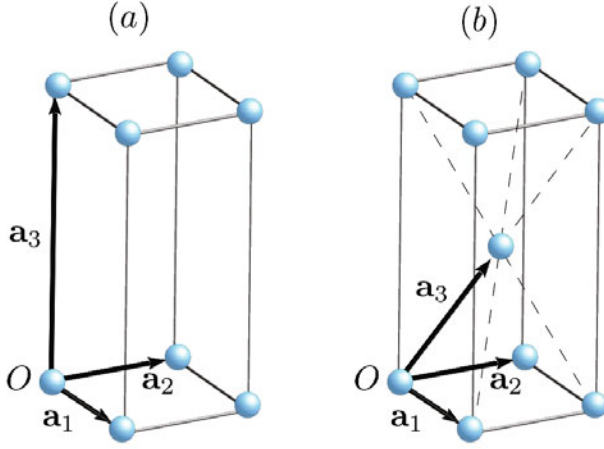


Fig. 1.13 Bravais lattices in 3D. Tetragonal system. The primitive cells of the two tetragonal lattices are depicted. (a) Simple tetragonal. (b) Body-centered tetragonal. A possible choice of the vectors \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 in the primitive cell is indicated

$$\mathbf{a}_1 = a\hat{\mathbf{x}} \quad ; \quad \mathbf{a}_2 = b\hat{\mathbf{y}} \quad ; \quad \mathbf{a}_3 = c\hat{\mathbf{z}}. \quad (1.11)$$

Base-centered ortonormal. It is obtained as in Fig. 1.14 (b). The primitive vectors are:

$$\mathbf{a}_1 = \frac{a}{2}\hat{\mathbf{x}} + \frac{b}{2}\hat{\mathbf{y}} \quad ; \quad \mathbf{a}_2 = b\hat{\mathbf{y}} \quad ; \quad \mathbf{a}_3 = c\hat{\mathbf{z}}. \quad (1.12)$$

Face-centered ortonormal. It is originated as in Figs. 1.14 (c). The primitive vectors are:

$$\mathbf{a}_1 = \frac{a}{2}\hat{\mathbf{x}} + \frac{b}{2}\hat{\mathbf{y}} \quad ; \quad \mathbf{a}_2 = \frac{a}{2}\hat{\mathbf{x}} + \frac{c}{2}\hat{\mathbf{z}} \quad ; \quad \mathbf{a}_3 = \frac{b}{2}\hat{\mathbf{y}} + \frac{c}{2}\hat{\mathbf{z}}. \quad (1.13)$$

Body-centered ortonormal. It is originated from the body-centered tetragonal lattice as depicted in Figs. 1.14 (d). The primitive vectors are:

$$\mathbf{a}_1 = a\hat{\mathbf{x}} \quad ; \quad \mathbf{a}_2 = b\hat{\mathbf{y}} \quad ; \quad \mathbf{a}_3 = \frac{a}{2}\hat{\mathbf{x}} + \frac{b}{2}\hat{\mathbf{y}} + \frac{c}{2}\hat{\mathbf{z}}. \quad (1.14)$$

Monoclinic system

The monoclinic system is originated from the ortonormal ones, after making the base to be a rhomb. Two different Bravais lattices are classified in the monoclinic system, simple and face-centered.

Simple monoclinic. It is depicted in Fig. 1.15 (a).

Base-centered monoclinic. It is depicted in Fig. 1.15 (b).

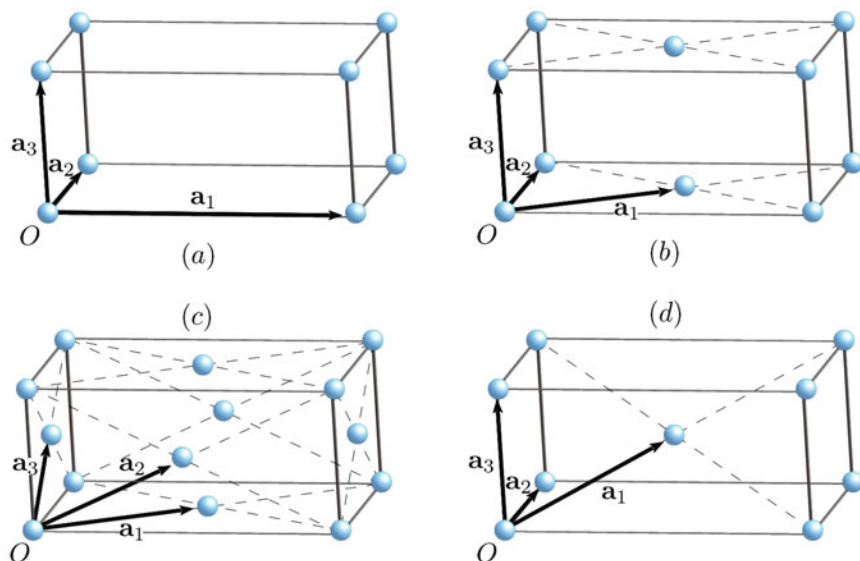


Fig. 1.14 Bravais lattices in 3D. Orthorhombic system. The primitive cells of the four orthorhombic lattices are depicted. (a) Simple orthorhombic. (b) Base-centered orthorhombic. (c) Face-centered orthorhombic. (d) Body-centered orthorhombic. A possible choice of the vectors \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 in the primitive cells is indicated

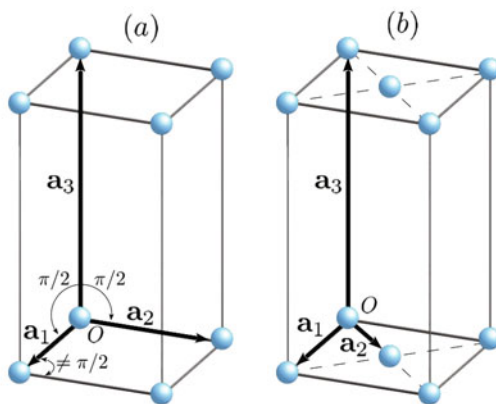
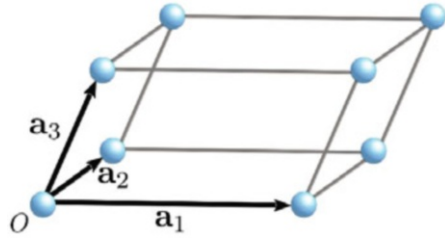


Fig. 1.15 Bravais lattices in 3D. Monoclinic system. The primitive cells of the two monoclinic lattices are depicted. (a) Simple monoclinic. (b) Base-centered monoclinic. A possible choice of the vectors \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 in the primitive cell is indicated

Triclinic system

The triclinic system is originated from the monoclinic one, after slanting vector \mathbf{a}_3 with respect to the direction orthogonal to the plane containing \mathbf{a}_1 and \mathbf{a}_2 . One simple triclinic lattice is classified here, as depicted in Fig. 1.16.

Fig. 1.16 Bravais lattices in 3D. Triclinic system. The primitive cell of the triclinic lattice is depicted. A possible choice of the vectors \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 in the primitive cell is indicated



Trigonal system

The trigonal system is originated from the simple cubic lattice, after deforming the cube along its diagonal: all the faces are rhombs, and each vertex is composed of three corners, each two of them forming the same angle. One simple trigonal lattice is classified here, as depicted in Fig. 1.17.

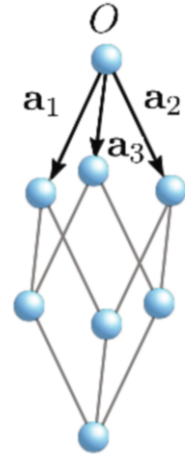


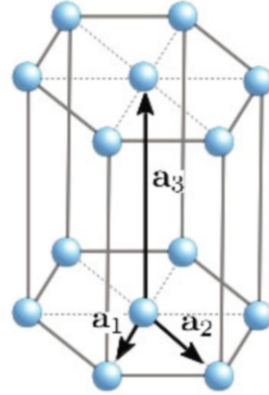
Fig. 1.17 Bravais lattices in 3D. Trigonal system. The primitive cell of the trigonal lattice is depicted. A possible choice of the vectors \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 in the primitive cell is indicated

Hexagonal system

The hexagonal system is obtained by layering hexagonal lattice planes one on top of the other, at distance c . The layering is arranged so that corresponding lattice points in adjacent planes are connected by lines perpendicular to the planes, as depicted in Fig. 1.18.

Considering the importance of graphite, as material which crystallizes in a hexagonal lattice, it is useful to explicitly consider the relevant details. The primitive vectors are:

Fig. 1.18 Bravais lattices in 3D. Hexagonal system. The primitive cell of the hexagonal lattice is depicted. A possible choice of the vectors \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 in the primitive cell is indicated



$$\mathbf{a}_1 = -\frac{a}{2}\hat{\mathbf{x}} + \frac{a\sqrt{3}}{2}\hat{\mathbf{y}} \quad ; \quad \mathbf{a}_2 = \frac{a}{2}\hat{\mathbf{x}} + \frac{a\sqrt{3}}{2}\hat{\mathbf{y}} \quad ; \quad \mathbf{a}_3 = c\hat{\mathbf{z}}, \quad (1.15)$$

where a is the size of the hexagon side. The cell volume amounts to $V_c = \sqrt{3}a^2c/2$. It turns out that $c > a$ in all materials related to the hexagonal lattice, and thus the number of first neighbors is 6.

A second type of hexagonal lattice may occur in real materials, which can be traced back to a simple hexagonal lattice with base, that is the so-called packed hexagonal lattice. In this structure, an additional lattice plane is layered between two adjacent lattice planes of the plain hexagonal structure, and located at mid-distance $c/2$ between them. The additional lattice plane is arranged so that the perpendicular line connecting two lattice sites belonging to the two original planes, falls into the barycenter of the middle plane, as depicted in Fig. 1.19. The relative orientation of the points in the middle lattice is different from that of the underlying plane. Thus, the packed hexagonal per se would not be a Bravais lattice. However, it can be described as a Bravais lattice with base vectors, the latter having coordinates $(0, 0, 0)$ and $(a/2, a/(2\sqrt{3}), c/2)$ from the origin. The relationship between a and c values can be derived from simple geometrical considerations, within an approximation where atoms behave as if they were hard spheres in contact with each other. As demonstrated in Problem 1.1, for the hexagonal close-packed structure one thus obtains $c/a = \sqrt{8/3} = 1.63299$.

Quick Questions

Q4. In a hexagonal lattice, consider a plane parallel to two adjacent lattice planes and located mid-way between them. Is a reflection operation with respect to this plane, a point-group symmetry operation of this lattice?

Answer. No, although it is a symmetry operation, it does not belong to the point group.

Q5. Consider a cubic lattice and the cube diagonal as a rotation axis.

Is this a symmetry axis with respect to rotation by a $2\pi/3$ angle?

Answer. Yes, it is. The cubic lattice point group contains rotational symmetry operations of $\pi/2$, $\pi/4$, $2\pi/3$, and parity inversions. The rotational operation of $2\pi/3$ around the cube diagonal passing through the origin performs the axes transformation $(x, y, z) \rightarrow (y, z, x)$. A similar reasoning applies to the other diagonals.

1.2.3 Nanocrystals

General considerations are given here on the structure of nanocrystals, useful in the rest of this textbook. A nanocrystal can be most simply viewed as a nanometric fragment of the corresponding macroscopic crystal. Alternatively, it can be viewed as a very large molecule. However, while in general a molecule has not a corresponding crystal on a macroscopic scale, a nanocrystal does have it, with relevant consequences. Along the same lines, a cluster of atoms is not to be considered a nanocrystal.

A huge variety of nanostructures has been synthesized, and an overall account of the wide existing literature is beyond the scope of this textbook. The focus is on nanocrystals confined in one-dimensional geometry such as carbon nanotubes, in two-dimensional such as graphene, and in three-dimensional, either spherical or elliptical, also known as quantum dots.

A finite number of atoms arranged on a chain is the simplest form of one-dimensional nanocrystal. At the beginning, such systems were considered for theoretical homework calculations of electronic states in a finite system, while current

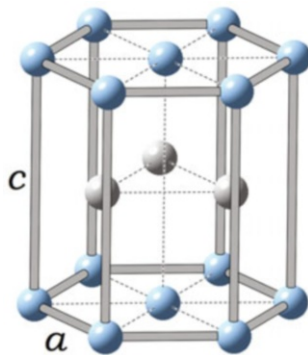


Fig. 1.19 Bravais lattices in 3D. The hexagonal close-packed structure is depicted

technologies now synthesize these structures in laboratory. One may predict that a wide use of these structures will occur in the next years to engineering nanodevices e.g. for electronic and biomedical applications. The simplicity of these systems makes them amenable to investigations, which turns out to be quite useful also for the comprehension of more complex systems. One such example is the study of the nature of electronic states and their evolution with increasing the system size up to those of the macroscopic crystal.

Let us consider one of the atomic the chain sketched in Fig. 1.20. The atom positions are evenly spaced along the line identified by the versor \hat{x} , so that $\mathbf{R}_n = na\hat{x}$ with $n = 1, 2, \dots, N$ and a the lattice spacing. The chain is embedded into a supporting macroscopic substrate. This is a general condition occurring in real nanostructures. The issue is in fact posed on the relevance of the possible interactions between the nanostructured system and its substrate. The substrate can be for example a silicon-based crystal, such as a SiO_2 matrix, or else a crystal deriving from the same atomic species than the chain. In general, nanostructure and substrate are to be considered together, since in most cases synthesis techniques entangle the two components. This implies a lot of details related to the complex underlying physical-chemistry, which are not deepened here.

As a general principle, if the whole chain and substrate are considered as one system, the electronic states in the chain are influenced by the interaction with the substrate atoms. If this were not the case, no adhesion of the chain to the substrate would occur. The issue here is not the existence of such an interaction, it being covalent, van der Waals or other, but how to identify the conditions under which

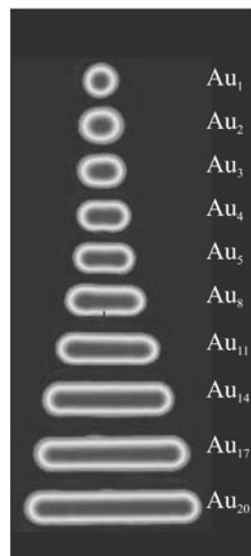


Fig. 1.20 One-dimensional (1D) nanocrystal structure. Au chains of different lengths on $\text{NiAl}(110)$, as recorded in Scanning Tunneling Microscopy (STM) topographic images [5]

the chain can be considered to be effectively isolated with respect to the electronic states.

To this aim, strategies can be adopted to engineer chains of atoms on a substrate, in a way that the energy range of the corresponding electronic states are well separated from each other. One such example is illustrated in Fig. 1.20, corresponding to the manipulation of single atoms on a surface by means of a Scanning Tunneling Microscopy (STM) tip. In a quantum mechanical perspective, the electronic wavefunctions would have a negligible overlap. Under these conditions nanostructure and substrate can be considered on a separate footing and it is possible to investigate how the electronic properties of the ideal infinite crystal, depend on the finite size of the nanostructure itself.

1.3 Lattice structure of specific crystals

This section is devoted to cast the ideal lattice structures illustrated in Sec. 1.2.2.2 into their actual realizations. Examples of elements crystallizing in either one of those lattice structures are illustrated in Sec. 1.3.1, and a similar illustration is performed in Sec. 1.3.2 for examples of compounds.

1.3.1 *Crystal structure of elements*

Concept

Most of chemical elements crystallizes into either a face-centered cubic or in a hexagonal close-packed lattice.

A qualitative two-step argument helps justifying this evidence, in a schematic representation where atoms are impenetrable hard spheres. In essence, one first notices that the face-centered cubic lattice is the structure that maximizes the space occupied by atoms, that is the ratio of the space occupied by atoms to the total available space or packing index, as elaborated in the quick question Q6.

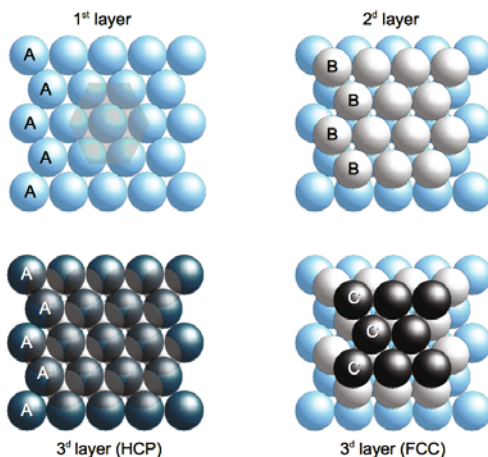
Definition

The packing fraction is defined as the ratio between the volume occupied by the ions in the unit cell, modeled as hard spheres, and the volume of the unit cell.

Then, one tackles the demonstration that face-centered cubic and hexagonal close-packed lattices share the same packing index. The construction depicted in Fig. 1.21 helps to see this result. Represent the atoms as identical hard spheres A,

Fig. 1.21 Face-centered cubic (fcc) and hexagonal close-packed (hcp) lattices share the same packing index.

The construction of the two lattices within the hard-sphere model is displayed. After the first layer of A atoms and the second one of B atoms are deposited to realize maximum packing, two different choices for the third layer lead to the same packing level but two different lattices: hcp in case ABA and fcc in case ABC



and consider a layer of such atoms that be in contact with each other. This arrangement leads to a hexagonal lattice, the sites being located at the sphere centers. Consider next a second layer of new atoms B posed on top of the first in a way to maximally fill in the available space: this goal can be reached by arranging the new atoms in the hollows originated by three in-touch spheres belonging to the first layer. Finally, consider a third layer of atoms C posed on top of the second, still in a way to maximally fill in the available space: this would be arranged with the new atoms C disposed so that their centers be located in correspondence either of the barycenter of three in-touch atoms of the first layer, or above the center of a sphere A in the first layer. By inspection, the structures that are obtained according to the above procedure are a face-centered cubic (fcc) lattice in the former case and a hexagonal close-packed (hcp) lattice in the latter.

Tables 1.1 and 1.2 list elements which crystallize into face-centered and body-centered cubic lattices, respectively, along with the experimental values of the related lattice parameter a . Table 1.3 instead, lists elements crystallizing into a hexagonal close-packed structure, together with the related experimental values of the lattice parameters a , c and resulting ratio c/a . It can be noticed that the experimental values of c/a are generally in good agreement with the theoretical prediction $\sqrt{8/3}$ derived from the hard-spheres model.

A significant remark here is the following: while lattice parameters are usually measured at low temperatures,

Concept

One and the same element may crystallize into different structures depending on external temperature and pressure conditions.

Table 1.1 Elements crystallizing into a fcc lattice and related values of the lattice parameter a . Data are referred form textbook [3]

| Element | a (Å) ^a | | Element | a (Å) ^a |
|---------|----------------------|--|---------|----------------------|
| Ne | 4.439 | | Ag | 4.086 |
| Ar | 5.256 | | Au | 4.078 |
| Kr | 5.721 | | Co | 3.548 |
| Xe | 6.197 | | Ni | 3.524 |
| Al | 4.050 | | Pd | 3.89 |
| Cu | 3.615 | | Pt | 3.923 |

^aExperimental values

Table 1.2 Elements crystallizing into a bcc lattice and related values of the lattice parameter a . Data are referred form textbook [3]

| Element | a (Å) ^a | | Element | a (Å) ^a |
|---------|----------------------|--|---------|----------------------|
| Li | 3.49 | | Cr | 2.88 |
| Na | 4.32 | | Nb | 3.300 |
| K | 5.25 | | Mo | 3.147 |
| Rb | 5.60 | | V | 3.024 |
| Cs | 6.07 | | W | 3.165 |

^aExperimental values

Table 1.3 Elements crystallizing into a hcp lattice and related values of lattice parameters a and c . Data are referred form textbook [3]

| Element | a (Å) ^a | c (Å) ^a | c/a^b |
|----------------|----------------------|----------------------|---------|
| He | 3.57 | 5.83 | 1.6331 |
| Be | 2.287 | 3.583 | 1.5667 |
| Mg | 3.209 | 5.210 | 1.6235 |
| Cs (β) | 3.98 | 6.52 | 1.6382 |
| Sr (β) | 4.32 | 7.06 | 1.6343 |
| Na | 3.767 | 6.154 | 1.6336 |
| Ni | 2.65 | 4.33 | 1.6340 |
| Ti | 2.95 | 4.69 | 1.5898 |
| Y | 3.647 | 5.731 | 1.5714 |
| Zn | 2.665 | 4.947 | 1.8562 |

^aExperimental values

^bNote that the theoretical value is $c/a = \sqrt{8/3} \simeq 1.6330$

For example, Fe has a face-centered cubic phase at high temperatures $T \simeq 1300$ K and a body-centered cubic structure at lower temperatures.

Several elements crystallize in structures corresponding to Bravais lattices with base. Considering the relevance of related materials, the structures occurring most often are usefully described in the following.

1.3.1.1 Elements crystallizing into Bravais lattices with base.

Diamond structure. One relevant example is displayed in Fig. 1.22 and is named after diamond. Besides diamond in fact, Silicon, Germanium, and gray (α -)Tin crystallize for example into this structure. Here, a face-centered cubic lattice is combined with a second one obtained from the first after a rigid translation by $(\hat{x} + \hat{y} + \hat{z})a/4$. The resulting structure can be described in terms of a face-centered cubic lattice with a base made up of two atoms located at $(0,0,0)$ and $(a/4, a/4, a/4)$. Overall, the atoms positions within the conventional cell are located at: $(0,0,0)$, $(a/2, a/2, 0)$, $(a/2, 0, a/2)$, $(0, a/2, a/2)$, $(a/4, a/4, a/4)$, $(3a/4, 3a/4, a/4)$, $(3a/4, a/4, 3a/4)$, and $(a/4, 3a/4, 3a/4)$.

As it is displayed in Fig. 1.22, each atom in the structure has four first neighbors. Thus, the four planes containing all possible threesomes of first-neighbor atoms form a tetrahedron. The first neighbors to the atom in $(0,0,0)$ are located at $(a/4, a/4, a/4)$, $(-a/4, a/4, -a/4)$, $(a/4, -a/4, -a/4)$, and $(-a/4, -a/4, a/4)$. The values of the lattice parameter a are 3.567 Å for C, 5.431 Å for Si, 5.657 Å for Ge, and 6.491 Å for gray-(α)Sn.

Observations on the mechanical properties of diamond bring about the opportunity to refine the hard-sphere approximation. It is known that diamond is a quite hard material. However, its packing index is much smaller than that of other materials crystallizing into cubic structures. Thus, additional considerations must be taken into account to explain this puzzle. It can be anticipated from Sec. 1.7 that these are related to the nature of chemical bonds: those keeping together the Carbon atoms are indeed covalent-like, and therefore highly directional. The tetrahedron structure reflects the typical configuration of the hybridized covalent bonds that each C atom shares with its first neighbors. These considerations manifest that chemical bonds

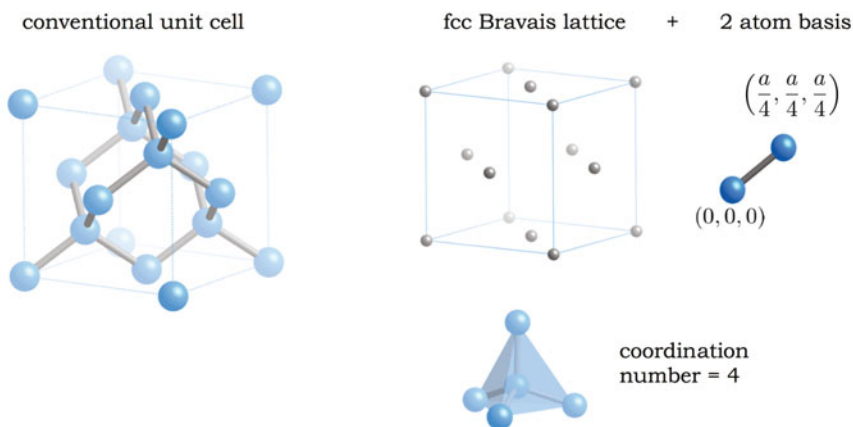


Fig. 1.22 Elements crystallizing into Bravais lattices with base: the diamond-like structure. Left panel: the conventional cell. Right panel: the non-Bravais lattice is described as a Bravais lattice with a two-atom base. The number of first neighbors, or coordination number, is also evidenced

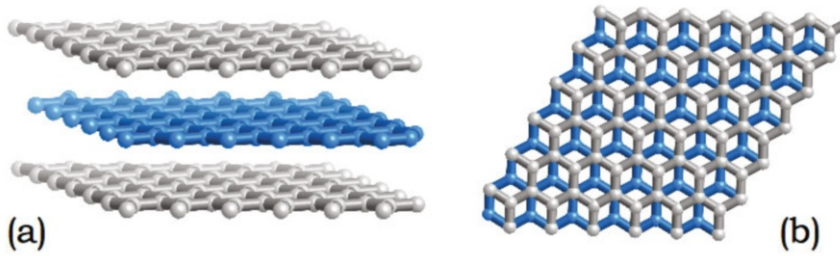


Fig. 1.23 Elements crystallizing into Bravais lattices with base: the graphite structure. (a) Lateral and (b) top view, showing that atoms in adjacent planes are staggered

play a relevant role in determining the physical properties of materials, besides crystal structure.

Graphite structure. Besides than in diamond structure, carbon can crystallize also as graphite. The graphite structure is illustrated in Fig. 1.23.

It is composed of single layers where atoms are arranged at the vertex of hexagons. Two characteristics are peculiar of graphite structure. First, as Fig. 1.23 (b) makes evident, atoms in adjacent planes are staggered: two corresponding hexagons in adjacent planes are each rotated by $\pi/3$ with respect to the other. Along these lines, the graphite structure can be described as a hexagonal Bravais lattice with a four-atoms base at positions $(0,0,0)$, $(0, a/\sqrt{3}, 0)$, $(0,0, c/2)$, and $(0, 2a/\sqrt{3}, c/2)$. This is displayed in the left and right panels of Fig. 1.24. Second,

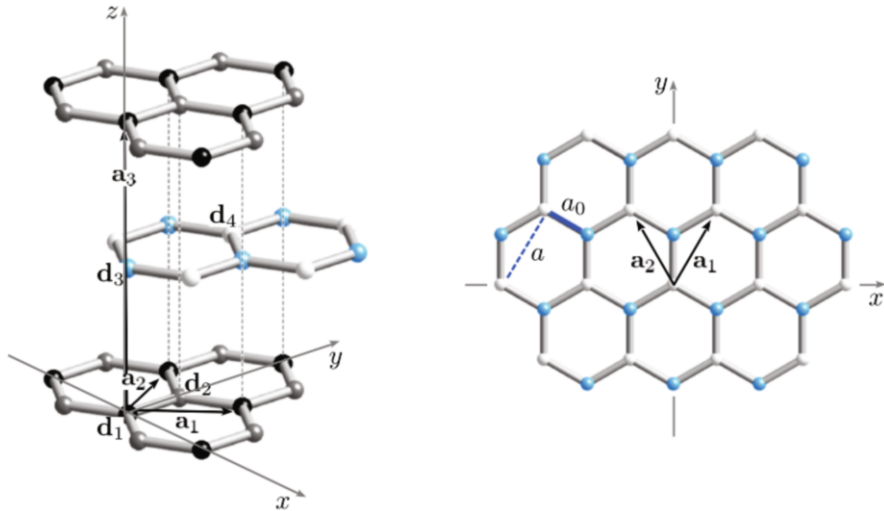


Fig. 1.24 Elements crystallizing into Bravais lattices with base: the graphite structure. Left panel: description in terms of a hexagonal Bravais lattice with a four atoms base. The primitive vectors \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 in (1.15) are indicated, along with the base vectors \mathbf{d}_1 , \mathbf{d}_2 , \mathbf{d}_3 , and \mathbf{d}_4 . Right panel: difference between the bond length a_0 of the hexagonal plane (solid blue line) and the size of the primitive translation vector $a = a_0\sqrt{3}$ (dashed blue line)

the lattice parameter within the plane amounts to 1.42 Å while the distance between adjacent planes is 3.35 Å. This is a manifestation of the nature of chemical bonds between the Carbon atoms, that is covalent within the plane and van der Waals-like in the direction perpendicular to them. This is indeed the reason why graphite easily tends to exfoliate. Once more, chemical bonds play a significant role in determining the mechanical properties of materials.

Quick Questions

Q6. Using the hard-sphere model, calculate the packing index I for simple, body-centered, and face-centered cubic lattices.

Answer. From simple geometrical considerations, one has in the three cases $I = \pi/6 = 0.524$, $I = \sqrt{3}\pi/8 = 0.680$, and $I = \pi/(3\sqrt{2}) = 0.740$. For example, the packing fraction of a bcc lattice can easily be calculated as follows. Within a hard sphere model where ions/atoms are assumed to be hard spheres with radius R , the conventional unit cell has a volume a^3 , with a the cubic lattice parameter. The conventional cell contains two non equivalent ions/atoms, one at the center of the cell, one at one of the vertexes, with volume $(4/3)\pi R^3$. By assuming the central atom in contact with its first neighbours, it is clear that the line of contact coincides with the cube diagonal and has a length $\sqrt{3}a = 4R$. Therefore, the packing fraction is

$$I = \frac{\text{volume occupied}}{\text{unit cell volume}} = \frac{2 \cdot \frac{4}{3}\pi R^3}{a^3} = \frac{2 \cdot \frac{4}{3}\pi R^3}{\left(\frac{4}{\sqrt{3}}R\right)^3} = 0.68 \quad (1.16)$$

Therefore, the closest packing in the bcc lattice fills 68% of the available space, and so on for the other structures.

Q7. Using the hard-sphere model, calculate the packing index of the diamond structure.

Answer. From simple geometrical considerations as in Q6, one has $I = \sqrt{3}\pi/16 \simeq 0.34$.

1.3.2 Crystal structure of compounds

The illustration is here limited to three significant and frequently occurring types of structures, into which two-elements compounds crystallize.

CsCl structure. Frequently occurring structures are those into which crystallize ionic salts of alkaline elements. A first type of such structures is named after CsCl and it is displayed in Fig. 1.25. Starting from a simple cubic lattice, this is obtained positioning either one of the two species, say Cs or Cl, at each cube vertex, while

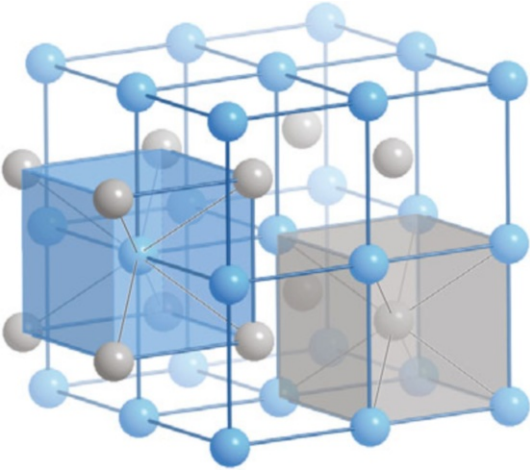


Fig. 1.25 Compounds crystallizing into the structure of CsCl. The structure is shown. It is evident that each ion of one species is surrounded by eight ions of the other species

at the center of the same cube is in turn positioned the other species, namely either Cl or Cs. Thus, the CsCl structure can be described as a simple cubic lattice with a two-atoms base. The two base atoms, one Cs and one Cl, are located at $(0,0,0)$ and $(a/2,a/2,a/2)$. The lattice parameter of CsCl is $a = 4.123 \text{ \AA}$. Other compounds crystallizing in this structure are listed in Table 1.4, together with the corresponding lattice parameters. Simple geometrical considerations help classifying the couples

Table 1.4 Compounds crystallizing into the CsCl structure and related values of the lattice parameter a (from [3])

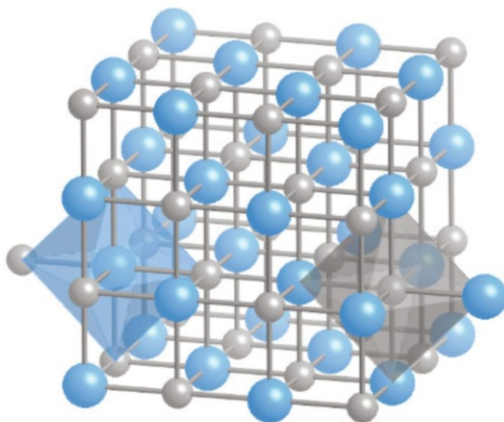
| Compound | $a \text{ (\AA)}^a$ | | Compound | $a \text{ (\AA)}^a$ |
|----------|---------------------|--|----------|---------------------|
| CsBr | 4.286 | | TlCl | 3.834 |
| CsI | 4.567 | | TlBr | 3.97 |
| RbCl | 3.74 | | TlI | 4.198 |
| ThTe | 3.827 | | | |

^a Experimental values

of ions which might crystallize into the CsCl structure, according to the size of their ionic radii R_1 and R_2 of, say Cs (smaller radius) and Cl (larger radius). Fig. 1.25 immediately shows that possible values for R_1 and R_2 are bounded by the conditions $R_1 + R_2 = \sqrt{3}a/2$ and $2R_2 \leq a$. As a result, one obtains $R_2/R_1 \leq 1/(\sqrt{3} - 1)$.

NaCl structure. Other salts of alkaline metals crystallize into a different structure, named after the sodium chloride, NaCl. This is displayed in Fig. 1.26. Starting from a face-centered cubic lattice, this is obtained by first positioning either one of the two species, say Na or Cl, at the lattice sites of a face-centered cubic Bravais lattice. Then, an atom of the other species, say Cl or Na, is positioned in between each pair of atoms of the other species spaced by a lattice distance a . Thus, the

Fig. 1.26 Compounds crystallizing into the structure of NaCl. The structure is shown. It is evident that each ion of one species is surrounded by six ions of the other species, positioned at the vertex of an octahedron



NaCl structure can be described as a face-centered cubic lattice with a two atoms base. The two base atoms in the primitive cell, one Na and one Cl, are located at $(0,0,0)$ and $(a/2,0,0)$. In the conventional cell instead, Na-like atoms are located at $(0,0,0)$, $(a/2,a/2,0)$, $(0,a/2,a/2)$, and $(a/2,0,a/2)$, while Cl-like atoms are located at $(a/2,0,0)$, $(0,a/2,0)$, $(0,0,a/2)$, and $(a/2,a/2,a/2)$.

Geometrical considerations similar to those discussed for CsCl lead to the conclusion that the ionic radii of ionic compounds crystallizing into this structure must satisfy the conditions $R_1 + R_2 = a/2$ and $2R_2 \leq \sqrt{2}a/2$, that is $R_2/R_1 \leq 1/(\sqrt{2} - 1)$. Other compounds crystallizing in the NaCl structure are listed in Table 1.5, together with the corresponding lattice parameters.

ZnS structure. A very common crystal structure taken by two-element compounds is named after the zincblend (ZnS), and it is displayed in Fig. 1.27. Starting from

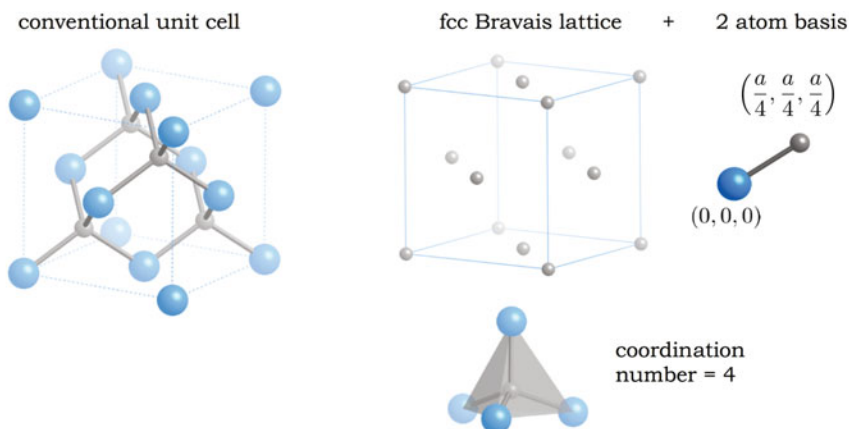


Fig. 1.27 Compounds crystallizing into the structure of ZnS (zincblende). The structure is shown. Left panel: the conventional cell. Right panel: the non-Bravais lattice is described as a fcc Bravais lattice with a two-atoms base. The number of first neighbors, or coordination number, is also evidenced

the diamond structure, this is obtained after locating each of the two species, say Zn and S, at the sites of each of the two intersecting face-centered cubic structures.

Similar geometrical considerations than for CsCl and NaCl structures, lead to the conditions $R_1 + R_2 = \sqrt{3}a/4$ and $2R_2 \leq \sqrt{2}a/2$, that is $R_2/R_1 \leq 1/\left[(3/2)^{\frac{1}{2}} - 1\right]$: the size of the two-species radii must satisfy the above consideration, in order to have a chance that the compound crystallizes in the ZnS structure. Other compounds crystallizing in the ZnS structure are listed in Table 1.6, together with the corresponding lattice parameters.

Quick Questions

Q8. Consider the parameters discussed up to now. Which among them would you think that the materials hardness depends on?

Answer. For what has been considered up to now, it depends at least on the packing index and on the nature of chemical bonds.

1.4 Reciprocal lattice

The conceptual and theoretical construction illustrated so far can be physically sound, but the question remains on how this knowledge has been built up, and in particular how the discussed crystal structures can be determined in an experiment. Several experiments deal with scattering of either particles or radiation and they therefore typically provide information primary on momentum and energy, with real space and time behaviors usually derived from actual observations. Thus, it is especially useful to introduce concepts and tools to describe crystal structures in momentum and energy space. The spatial periodicity of Bravais lattices makes a formulation in momentum space particularly handy, since the lattice-related physical quantities can be represented by means of Fourier-series expansions. To this aim, the concept of reciprocal lattice comes at hand.

Concept

Reciprocal lattice and related tools are quite useful to handle crystal structures in momentum space, especially to assist the comprehension of experiments and improve the ability for their theoretical interpretation and experimental design.

The main concepts are introduced in this section, where the related tools are also developed for theoretical calculations. In this backward pathway, the acquired tools are then exploited in Sec. 1.6, to understand the experimental issues.

1.4.1 Definition of reciprocal lattice

In order to define the reciprocal lattice, the first step is to identify the vectors, say \mathbf{G} , corresponding in momentum space to real lattice vectors \mathbf{R} , (1.3). To this aim, consider the plane wave $e^{i\mathbf{G}\cdot\mathbf{r}}$. Let us look for all those vectors \mathbf{G} , able to tune this plane wave to be periodical with respect to translations by a lattice vector \mathbf{R} , that is:

$$e^{i\mathbf{G}\cdot(\mathbf{r}+\mathbf{R})} = e^{i\mathbf{G}\cdot\mathbf{r}} \quad \forall \mathbf{R}. \quad (1.17)$$

Expression (1.17) is equivalent to

$$e^{i\mathbf{G}\cdot\mathbf{R}} = 1 \quad \forall \mathbf{R}. \quad (1.18)$$

Now, it turns out that the \mathbf{G} vectors satisfying (1.18) compose a lattice, whose elementary translation vectors are:

Definition

$$\begin{aligned} \mathbf{b}_1 &= 2\pi \frac{\mathbf{a}_2 \wedge \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \wedge \mathbf{a}_3)}, \\ \mathbf{b}_2 &= 2\pi \frac{\mathbf{a}_3 \wedge \mathbf{a}_1}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \wedge \mathbf{a}_3)}, \\ \mathbf{b}_3 &= 2\pi \frac{\mathbf{a}_1 \wedge \mathbf{a}_2}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \wedge \mathbf{a}_3)}. \end{aligned} \quad (1.19)$$

Let us prove the statement. Since $\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi \delta_{i,j}$, it easily verified that the vectors

Table 1.5 Compounds crystallizing into the NaCl structure and related values of the lattice parameter a (from [3])

| Compound | a (Å) ^a | | Compound | a (Å) ^a | | Compound | a (Å) ^a |
|----------|----------------------|--|----------|----------------------|--|----------|----------------------|
| LiH | 4.085 | | AgCl | 5.547 | | RbH | 6.037 |
| LiF | 4.027 | | AgBr | 5.775 | | RbF | 5.64 |
| LiCl | 5.120 | | MgO | 4.211 | | RbCl | 6.58 |
| LiBr | 5.501 | | CaO | 4.81 | | RbBr | 6.85 |
| LiI | 6.000 | | CdO | 4.695 | | RbI | 7.342 |
| KH | 5.70 | | NaH | 4.88 | | MnO | 4.445 |
| KF | 5.347 | | NaF | 4.620 | | FeO | 4.31 |
| KCl | 6.293 | | NaCl | 5.63 | | NiO | 4.168 |
| KBr | 6.60 | | NaBr | 5.973 | | TiO | 4.177 |
| AgF | 4.92 | | NaI | 6.473 | | MnS | 5.224 |

^a Experimental values

Table 1.6 Compounds crystallizing into the ZnS (zincblend) structure and related values of the lattice parameter a (from [3])

| Compound | a (Å) ^a | | Compound | a (Å) ^a |
|----------|----------------------|--|----------|----------------------|
| BN | 3.615 | | AlP | 5.451 |
| BP | 4.538 | | AlAs | 5.62 |
| BAs | 4.777 | | AlSb | 6.1347 |
| GaP | 5.4505 | | InP | 5.8687 |
| GaAs | 5.6537 | | InAs | 6.036 |
| GaSb | 6.118 | | InSb | 6.4782 |
| CdS | 5.818 | | HgS | 5.8517 |
| CdTe | 6.480 | | HgTe | 6.429 |
| CuF | 4.255 | | ZnS | 5.4093 |
| CuCl | 5.4057 | | ZnSe | 5.6676 |
| CuBr | 5.6905 | | ZnTe | 6.089 |
| CuI | 6.0427 | | | |

^a Experimental values

$$\mathbf{G} = r_1 \mathbf{b}_1 + r_2 \mathbf{b}_2 + r_3 \mathbf{b}_3 \quad (1.20)$$

with r_1 , r_2 , and r_3 relative integer numbers have the property

$$\mathbf{G} \cdot \mathbf{R} = 2\pi(n_1 r_1 + n_2 r_2 + n_3 r_3), \quad (1.21)$$

and thus satisfy (1.18). In conclusion, all \mathbf{G} of the form (1.20) are the vectors in momentum space which correspond to the real-space or direct lattice vectors \mathbf{R} . It remains now to prove that they are unique. Indeed, if a vector \mathbf{G} exists in a form different from (1.20), this should anyway be expressed as a superposition of \mathbf{b}_i but with non-integer coefficients r_i , with $i = 1, 2$, and 3 . But if this were the case, it would always be possible to find from (1.21) integer values n_1 , n_2 , and n_3 for which $e^{i\mathbf{k} \cdot \mathbf{R}} \neq 1$, thereby contradicting (1.18).

1.4.2 Properties of the reciprocal lattice

Definitions (1.18)–(1.20) lead to the following significant consequential properties:

Properties

P1. The volume size V_r of the reciprocal-lattice elementary cell is $V_r = |\mathbf{b}_1 \cdot (\mathbf{b}_2 \wedge \mathbf{b}_3)|$. From (1.19), this amounts to be $V_r = (2\pi)^3 / V_c$. Thus, the product of cell volumes of the direct and reciprocal lattices is a constant, namely the volumes are complementary. This is to be expected as a general property of Fourier-transform operations.

P2. The set of functions $e^{i\mathbf{G}\cdot\mathbf{r}}$ labeled by different \mathbf{G} vectors is orthogonal inside the elementary cell of the direct lattice. Indeed:

$$\int_{V_c} e^{i(\mathbf{G}-\mathbf{G}')\cdot\mathbf{r}} d\mathbf{r} = \int_{V_c} e^{i\mathbf{G}''\cdot\mathbf{r}} d\mathbf{r}, \quad (1.22)$$

with \mathbf{G} , \mathbf{G}' , and \mathbf{G}'' reciprocal lattice vectors. Inserting $\mathbf{r} = u\mathbf{a}_1 + v\mathbf{a}_2 + w\mathbf{a}_3$ with u, v , and $w \in [0, 1)$, and using \mathbf{G} from (1.20) one obtains:

$$\int_{V_c} e^{i\mathbf{G}''\cdot\mathbf{r}} d\mathbf{r} = V_c \int_0^1 du \int_0^1 dv \int_0^1 dw e^{i2\pi(r_1 u + r_2 v + r_3 w)}. \quad (1.23)$$

Here, V_c is the Jacobian resulting from the change of variables $(x, y, z) \rightarrow (u, v, w)$. The integral (1.23) is non vanishing if and only if $r_1 = r_2 = r_3 = 0$.

P3. The set of functions $e^{i\mathbf{G}\cdot\mathbf{r}}$ is also complete, as resulting from the Fourier expansion of periodical functions in the direct lattice. In essence, if $f(\mathbf{r}) = f(\mathbf{r} + \mathbf{R}) \forall \mathbf{R}$, then $f(\mathbf{r})$ can be Fourier expanded as

$$f(\mathbf{r}) = \sum_{\mathbf{G}} C_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} \quad (1.24)$$

with

$$C_{\mathbf{G}} = \frac{1}{V_c} \int_{V_c} e^{-i\mathbf{G}\cdot\mathbf{r}} f(\mathbf{r}) d\mathbf{r}. \quad (1.25)$$

According to the theory of Fourier series, the convergence is punctual, uniform or on average, depending on $f(\mathbf{r})$ properties.

P4. Nicely enough, the reciprocal of a reciprocal lattice is the direct lattice. This is easily demonstrated applying one more time the construction (1.19) to the \mathbf{b} vectors in place of the \mathbf{a} . As an immediate consequence, whatever function $f(\mathbf{k})$ periodic on the reciprocal lattice, namely $f(\mathbf{k}) = f(\mathbf{k} + \mathbf{G}) \forall \mathbf{G}$, can be Fourier expanded according to

$$f(\mathbf{k}) = \sum_{\mathbf{R}} C_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \quad \text{with} \quad C_{\mathbf{R}} = \frac{1}{V_r} \int_{V_r} e^{-i\mathbf{k}\cdot\mathbf{R}} f(\mathbf{k}) d\mathbf{k}. \quad (1.26)$$

Definition

The Wigner-Seitz cell of the reciprocal lattice is named Brillouin zone.

The geometrical construction of the reciprocal lattice of a Bravais lattice along with its Brillouin zone is performed according to the following

Procedure

Step 1. Identify the Bravais lattice and the corresponding expressions for \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 , as derived in Sec. 1.2.2.2.

Step 2. Apply the construction (1.19), after inserting the \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 and carry out the calculation.

Examples

Here follow a few examples, where this procedure is repeatedly applied.

Simple cubic lattice. One obtains:

$$\mathbf{b}_1 = \frac{2\pi}{a} \hat{\mathbf{x}} \quad ; \quad \mathbf{b}_2 = \frac{2\pi}{a} \hat{\mathbf{y}} \quad ; \quad \mathbf{b}_3 = \frac{2\pi}{a} \hat{\mathbf{z}}. \quad (1.27)$$

Thus, the reciprocal lattice of a simple cubic direct lattice with side a is simple cubic lattice with side $2\pi/a$.

Body-centered cubic lattice. One obtains:

$$\mathbf{b}_1 = \frac{4\pi}{a} \frac{1}{2} (\hat{\mathbf{x}} + \hat{\mathbf{y}}) \quad ; \quad \mathbf{b}_2 = \frac{4\pi}{a} \frac{1}{2} (\hat{\mathbf{y}} + \hat{\mathbf{z}}) \quad ; \quad \mathbf{b}_3 = \frac{4\pi}{a} \frac{1}{2} (\hat{\mathbf{x}} + \hat{\mathbf{z}}). \quad (1.28)$$

Thus, the reciprocal lattice of a body-centered cubic lattice with side a is a face-centered cubic lattice with side $4\pi/a$. The corresponding Brillouin zone is displayed in Fig. 1.28.

Face-centered cubic lattice. One obtains:

$$\mathbf{b}_1 = \frac{4\pi}{a} \frac{1}{2} (\hat{\mathbf{x}} + \hat{\mathbf{y}} - \hat{\mathbf{z}}) \quad ; \quad \mathbf{b}_2 = \frac{4\pi}{a} \frac{1}{2} (-\hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}}) \quad ; \quad \mathbf{b}_3 = \frac{4\pi}{a} \frac{1}{2} (\hat{\mathbf{x}} - \hat{\mathbf{y}} + \hat{\mathbf{z}}). \quad (1.29)$$

Thus, the reciprocal conventional cell of a face-centered cubic direct-lattice conventional cell with side a , is a body-centered cubic lattice with side $4\pi/a$. The corresponding Brillouin zone is displayed in Fig. 1.29.

Simple tetragonal lattice. One obtains:

$$\mathbf{b}_1 = \frac{2\pi}{a} \hat{\mathbf{x}} \quad ; \quad \mathbf{b}_2 = \frac{2\pi}{a} \hat{\mathbf{y}} \quad ; \quad \mathbf{b}_3 = \frac{2\pi}{c} \hat{\mathbf{z}}. \quad (1.30)$$

Thus, the reciprocal lattice of a simple tetragonal direct lattice with parameters a and c , is a simple tetragonal lattice with two sides amounting to $2\pi/a$ and the third to $2\pi/c$.

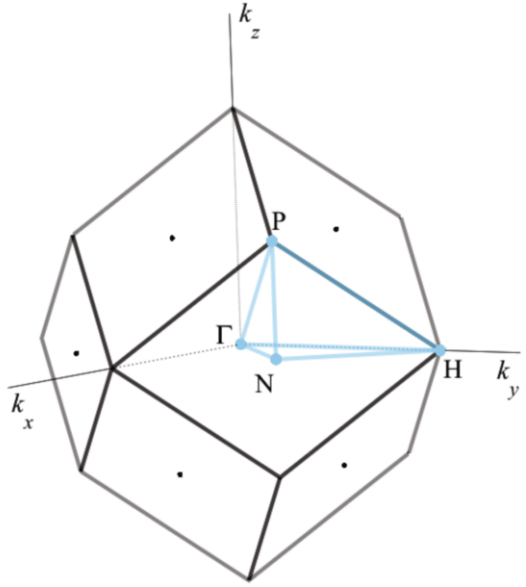


Fig. 1.28 Brillouin zone (BZ) of reciprocal lattices. The BZ of a body-centered cubic direct lattice

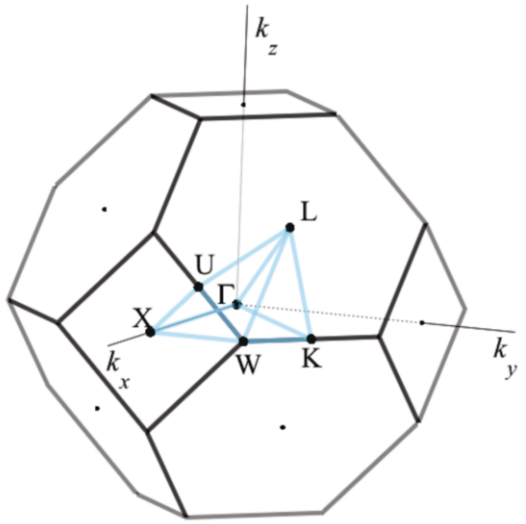


Fig. 1.29 Brillouin zone (BZ) of reciprocal lattices. The BZ of a face-centered cubic direct lattice

The tetragonal lattice embodies to some approximation the case of one-dimensional lattices. A 1D lattice can indeed be obtained as the limiting case of a simple tetragonal lattice, while either $a \rightarrow 0$ or $a \rightarrow \infty$. In the former case $\mathbf{b}_1 \rightarrow \infty$ and $\mathbf{b}_2 \rightarrow \infty$, while in the latter $\mathbf{b}_1 \rightarrow 0$ and $\mathbf{b}_2 \rightarrow 0$. However, in both cases the limit $b_3 \rightarrow 2\pi/c$ holds, representing a system that is invariant for arbitrary translations in the xy -plane, and periodic with period c along the z -axis. That is, an effectively 1D crystal. After performing the same limiting operations, (1.30) therefore provides the

Brillouin zone of a one-dimensional lattice with lattice parameter c . This results to be the set of k -vectors in the segment $-\pi/c \leq k < \pi/c$.

Hexagonal lattice. One obtains:

$$\mathbf{b}_1 = \frac{2\pi}{a} \left(\hat{\mathbf{x}} + \frac{1}{\sqrt{3}} \hat{\mathbf{y}} \right) ; \quad \mathbf{b}_2 = \frac{2\pi}{a} \left(-\hat{\mathbf{x}} + \frac{1}{\sqrt{3}} \hat{\mathbf{y}} \right) ; \quad \mathbf{b}_3 = \frac{2\pi}{c} \hat{\mathbf{z}}. \quad (1.31)$$

Thus, the reciprocal lattice of a hexagonal direct lattice is a hexagonal lattice. The corresponding Brillouin zone is displayed in Fig. 1.30.

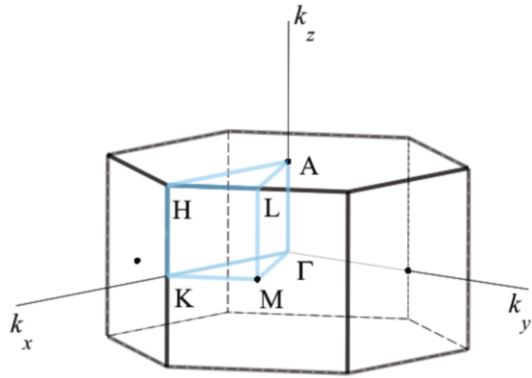


Fig. 1.30 Brillouin zone (BZ) of reciprocal lattices. The BZ of a hexagonal direct lattice

From the definition of Brillouin zone, it immediately descends the following

Properties

An arbitrary vector \mathbf{q} in reciprocal space can be always written as

$$\mathbf{q} = \mathbf{k} + \mathbf{G}, \quad (1.32)$$

where \mathbf{k} lies inside the first Brillouin zone and \mathbf{G} is a reciprocal lattice vector.

1.5 Lattice planes

Each representation of crystal structures brings about its own convenience. Theoretical calculations and handling of experimental data often require the ability to rapidly envision the crystal structure and especially to switch among different representations. This section is devoted to develop a useful switching tool in the form of a theorem, also leading to an alternative labeling of lattice planes.

1.5.1 An useful theorem

In essence, the theorem states that selected families of planes in the direct lattice can be easily labeled by reciprocal-lattice vectors. By definition,

Definition

A family of lattice planes is a set of evenly-spaced parallel planes containing lattice points, in a way that each lattice point belongs to a plane of the family.

In detail, the theorem states that

Theorem 1.1. a) *Given a family of lattice planes evenly-spaced by a distance d , $\mathbf{G} = 2\pi\mathbf{n}/d$ is a vector of the reciprocal lattice, with \mathbf{n} being the unit vector identifying the direction perpendicular to the planes. Precisely, among all possible vectors perpendicular to the planes, \mathbf{G} is that with smallest size.*

b) *Conversely, given a vector \mathbf{G} of the reciprocal lattice, and \mathbf{G}_0 the vector parallel to \mathbf{G} and having the smallest size, there exists a family of lattice planes perpendicular to \mathbf{G}_0 , which are evenly spaced by $d = 2\pi/G_0$.*

Concept

In conclusion, at any given direction a one-to-one correspondence exists between families of lattice planes and vectors of the reciprocal lattice with smallest size.

Proof. (a) One has first to express the condition according to which \mathbf{G} be perpendicular to the planes of the family. If \mathbf{r} is the vector identifying the points of one of those planes, the condition can be expressed as $\mathbf{G} \cdot \mathbf{r} = A$, with A a constant. The vectors \mathbf{r} and \mathbf{r}' singling out points belonging to a subsequent plane are related to each other by $\mathbf{r}' = \mathbf{r} + \mathbf{n}d$. On the other hand, translational invariance implies that $e^{i\mathbf{G} \cdot (\mathbf{r} + \mathbf{n}d)} = e^{i\mathbf{G} \cdot \mathbf{r}} = A'$ with A' a new constant. Since one of the family planes contains the point $\mathbf{R} = 0$, it turns out that $A' = 1$. Therefore, $\mathbf{G} = 2\pi\mathbf{n}/d$ is by definition a vector of the reciprocal lattice. This is also unique. Indeed, if a reciprocal vector \mathbf{G}' exist that is parallel to but different from \mathbf{G} , and such that $|\mathbf{G}'| < |\mathbf{G}|$, this would imply that the planes be spaced by an amount $d' > d$. The resulting family of planes would not contain all the lattice points, contrary to the original definition.

(b) The condition $\mathbf{G}_0 \cdot \mathbf{r} = A$ with A a constant, defines a plane perpendicular to \mathbf{G}_0 . Thus, following the proof given in (a) above, the condition $e^{i\mathbf{G}_0 \cdot \mathbf{r}} = 1$ defines a family of planes parallel to each other and evenly-spaced by $d = 2\pi/G_0$. All the lattice points belong to these planes and no plane exists which does not contain lattice points.

1.5.2 Miller indexes

The one-to-one correspondence between families of lattice planes and vectors of the reciprocal lattice provides a quite short-hand labeling of the lattice planes. This is given by three numbers r_1 , r_2 , and r_3 , identifying the least sized vector of the reciprocal lattice perpendicular to the family. In particular,

Definition

The integer numbers r_1 , r_2 , and r_3 with no common factors, which identify $\mathbf{G} = r_1\mathbf{b}_1 + r_2\mathbf{b}_2 + r_3\mathbf{b}_3$ and thus the family of lattice planes perpendicular to it, are named Miller indexes.

Miller indexes have a simple geometrical meaning. A plane of the family, represented by the equation $\mathbf{G} \cdot \mathbf{r} = A$, intersects the axis \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 into the points $x\mathbf{a}_1$, $y\mathbf{a}_2$, and $z\mathbf{a}_3$, where the coordinates $x = A/(2\pi r_1)$, $y = A/(2\pi r_2)$, $z = A/(2\pi r_3)$ contain the Miller indexes.

The identification of the family lattice planes from Miller indexes (r_1, r_2, r_3) (and viceversa) is accomplished by the following

Procedure

Step 1. Write the expression for \mathbf{G}_0 , that is $\mathbf{G}_0 = r_1\mathbf{b}_1 + r_2\mathbf{b}_2 + r_3\mathbf{b}_3$, along with its modulus providing the distance $d = 2\pi/G_0$ between adjacent lattice planes.

Step 2. Write the expression locating the lattice points in a plane of the selected family, that is $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$.

Step 3. Represent a plane of the family by means of its defining condition $\mathbf{R} \cdot \mathbf{G}_0 = A$ with A a constant.

Step 4. Explicitly calculate $\mathbf{R} \cdot \mathbf{G}_0$, to obtain a relation connecting the integer numbers n_1 , n_2 , and n_3 . This relation provides the infinite solutions for n_1 , n_2 , and n_3 identifying the family of planes.

Examples

Consider a face-centered cubic lattice and a family of lattice planes with Miller indexes $(1,0,0)$. The equation representing the family is determined following the above procedure. One first writes the expression for $\mathbf{G}_0 = \mathbf{b}_1$, its modulus $G_0 = 2\sqrt{3}\pi/a$, and the distance between to adjacent planes $d = 2\pi/G_0 = a/\sqrt{3}$. The lattice points satisfy the condition $\mathbf{R} \cdot \mathbf{G}_0 = A$ with A an arbitrary constant. Thus, one calculates $\mathbf{R} \cdot \mathbf{G}_0 = 2\pi n_1$. After fixing n_1 , all the lattice points of the selected

plane are identified by varying n_2 and n_3 within the ensemble of the relative integer numbers.

Quick Questions

Q9. Identify the vector \mathbf{G}_0 expressing the family of planes with Miller indexes $(1, 1, 0)$ in a simple cubic lattice.

Answer. One has $\mathbf{G}_0 = \mathbf{b}_1 + \mathbf{b}_2$, and thus $G_0 = 2\pi\sqrt{2}/a$ and $d = 2\pi/G_0 = a/\sqrt{2}$. A lattice plane of the family is represented by the relation $\mathbf{R} \cdot \mathbf{G}_0 = A$ with A a constant, that is $\mathbf{R} \cdot \mathbf{G}_0 = 2\pi(n_1 + n_2)$. This implies that n_1 and n_2 be such that $n_1 + n_2 = n$ with n given and n_3 arbitrary.

Q10. Identify the vector \mathbf{G}_0 expressing the family of planes with Miller indexes $(4, 2, 2)$ in a simple cubic lattice.

Answer. In a way similar to Q9 above, one has $\mathbf{G}_0 = 2\mathbf{b}_1 + \mathbf{b}_2 + \mathbf{b}_3$, $G_0 = 2\pi\sqrt{6}/a$, and $d = 2\pi/G_0 = a/\sqrt{6}$. Thus $\mathbf{R} \cdot \mathbf{G}_0 = 2\pi(2n_1 + n_2 + n_3)$, leading to $2n_1 + n_2 + n_3 = n$ with n given.

1.6 Experimental determination of system structures: Scattering of radiation and particles from crystals

1.6.1 Introduction

This section is devoted to apply the concepts and tools developed so far to the experimental determination of crystal structures. On general grounds,

Concept

Investigation of a system structure generally requires particle- or radiation-like probes able to couple to the observable under study, and acting on a length scale of the order of the system size or smaller.

In turn, the typical radiation frequencies $\nu = c/\lambda$ and thus energies $E = h\nu$, with c the speed of light, inversely scale with the wavelength λ . In the case of particles, involved energies are of the order of $E = \hbar^2/(2m\lambda_{dB}^2)$, with m the particle mass, inversely scaling with the squared de Broglie wavelength λ_{dB} of the associated wave.

Standard experiments are set by shooting the target with particles or radiation having selected energy E_{in} , wavenumber \mathbf{k}_{in} , and intensity I_0 . The probe is scattered out of the target, in principle along different directions, with new energy E_f ,

wavenumber \mathbf{k}_f , and intensity I_f . The measured quantities are the differences in energy $\Delta E = E_f - E_{in}$, wavenumber $\Delta \mathbf{k} = \mathbf{k}_{in} - \mathbf{k}_f$, and the intensity I_f related to the incident one I_0 : from these quantities the system structure is inferred.

The source of the probing radiation beam is placed far away from the target, at distance L much larger than the typical target size d ($d \ll L$). Under these conditions, the beam can be described in terms of plane waves, introducing a very useful simplification. The detector is for similar reasons placed at distance L' from the target, such that $d \ll L'$. In this limit, close to the detector each field scattered from single parts of the target, can be described as a plane wave with wavevectors parallel to each other.

Concept

Experiments where $\Delta E = 0$ correspond to elastic scattering: they allow the determination of the geometrical structure of the target. Experiments where $\Delta E \neq 0$ correspond to inelastic scattering: they also allow the determination of excitation energies of the target system.

In the following, more detailed conditions are discussed that are useful to set and observe scattering of radiation or particles from crystals. Calculations suited to interpret experimental data are then provided for selected examples. Finally, experimental techniques are discussed, to determine the structure of crystal and nanosystems.

1.6.2 Scattering of radiation and particles from crystals: basic facts

When applied to the determination of atomic densities in crystal structures, the simple concepts enlightened in the introduction lead to the conclusion that λ is to be of the order of the lattice spacing, $\lambda \simeq 10^{-8}$ cm. Thus, either X-rays are needed, or electrons with energies on the scale of 100 eV, or else neutrons with energies of the order of 10^{-1} eV. The difference in typical energies between electron and neutron probes is due to their different masses. On the other hand, use of X-rays is not convenient for inelastic scattering experiments, since X-ray incoming photons have energies E_{in} in the range 1–10 KeV and typical $\Delta E = E_f - E_{in}$ values would be quite small relative to E_{in} . Scattering from particles, electrons or neutrons, which have much lower incoming energies, are better suited to inelastic experiments. Besides, one would better use neutrons to investigate the crystal bulk and electrons for surface studies, since lighter charged particles experience stronger interactions with the medium and thus have short penetration depths. Progresses in quantum physics have made available electrons and neutrons probes for structure, electronic, and vibrational excitations studies.

This section treats only the X-ray elastic scattering case with $\Delta E = 0$, suited to determine the geometrical structure of crystals. Inelastic scattering turns to be useful in Chap. 5 where excitations are investigated. The first X-ray experiment was per-

formed in 1913 by von Laue, who has at one time demonstrated the electromagnetic nature of X-rays and the regular arrangement of atoms in crystal structures.

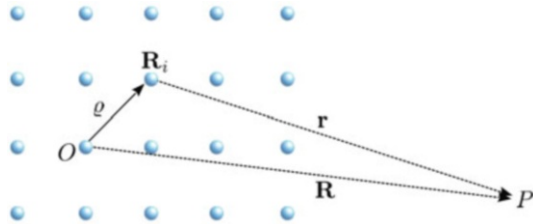
The large source-target-detector distances on the scale of the sample size, discussed to be best suited for experimental conditions, would suggest that the crystal sample be as smallest as possible. In fact, this has not to be the case. As already discussed at the beginning of this Chapter, a finite crystal might mimic well all the properties of an infinite crystal if at least two conditions are satisfied: the number of atoms on its surface has to be much smaller than their total number, and the size of the crystal must be much larger than typical ranges of the atomic interactions in the bulk.

Concept

It is known that a charged particle in an electromagnetic field acts as an antenna. That is, it is accelerated by the electromagnetic force, becoming in turn an emitting source of radiation having the same frequency of the incoming field.

Referring to Fig. 1.31, let P be an observation point, external to the crystal, at position \mathbf{R} with respect to a fixed origin. The incoming radiation is scattered along a direction perpendicular to the vector connecting source and detector, laying into the plane identified by \mathbf{r} and the vector representing the charge acceleration.

Fig. 1.31 Elastic scattering of radiation from an atom. A sketch of the experimental scheme is given. The atom is located at position R_i and P is a point at the detector



Q_ρ located inside the target crystal at position ρ with respect to the origin O , becomes a point source of spherical waves. This means that it emits an electric field, which in turn can be expressed as:

$$E_\rho \propto Q_\rho \frac{e^{ikr}}{r} e^{i(\mathbf{k} \cdot \rho - \omega t)}, \quad (1.33)$$

where r is the distance between charge and detector and \mathbf{k} is the wavenumber of the incoming radiation with frequency ω . The quantity r can be expressed in terms of the distances R between reference atom and detector, and ρ between reference atom and charge, that is:

$$r^2 = R^2 + \rho^2 - 2R\rho \cos \theta,$$

where θ is the angle between ρ and \mathbf{R} . Thus, one obtains to first order in ρ/R and under the given conditions $\rho \ll R$:

$$r = R \left(1 - \frac{\rho}{R} \cos \theta \right) = R - \rho \cos \theta . \quad (1.34)$$

Equation (1.34) has a crucial physical meaning: the phase kr of the field e^{ikr} emitted by the charge in $\boldsymbol{\rho}$ is shifted by the amount $-k\rho \cos(\theta)$ with respect to the reference phase kR of the charge at the origin. In the $1/r$ Taylor expansion of (1.33), only the term in $1/R$ gives rise to a radiative field. Considering that the detector is very far, vectors \mathbf{R} and \mathbf{r} can be approximated to be almost parallel to each other, so that the corresponding electric fields at the detector share the same direction. Thus, the total field at the detector results from summing the contributions from all the charges and can be approximated to be:

$$E \propto \left(\sum_{\boldsymbol{\rho}} \frac{Q_{\boldsymbol{\rho}}}{R} e^{-ik\rho \cos(\theta)} e^{i\mathbf{k} \cdot \boldsymbol{\rho}} \right) e^{i(kR - \omega t)} . \quad (1.35)$$

The phase $k\rho \cos(\theta)$ can be written as $\mathbf{k}' \cdot \boldsymbol{\rho}$, where \mathbf{k}' is the wavevector of modulus k directed along \mathbf{R} . Equation (1.35) can be expressed as

$$E \propto \sum_{\boldsymbol{\rho}} Q_{\boldsymbol{\rho}} e^{i\Delta \mathbf{k} \cdot \boldsymbol{\rho}} , \quad (1.36)$$

where $\Delta \mathbf{k} = \mathbf{k} - \mathbf{k}'$. Consider that the immaterial phase factor $e^{i(kR - \omega t)}$ and the term $1/R$ have been omitted, since they do not depend on the charge distribution of the target.

Formula (1.36) describes the elastic scattering of electromagnetic radiation from a general system of charges which are not necessarily arranged in a perfect crystal. The actual calculation of the electric field can be performed after specifying the atom species. One more simplification can be introduced here by neglecting the contribution of the scattering from nuclei: their acceleration is inversely proportional to the mass in the present classical view, and therefore it is much smaller than that of the electrons, since the ratio m/M of the electron m to the nucleus M masses is $m/M \ll 1$.

1.6.3 Calculation of scattered intensities

The actual calculation of the sum in (1.36) for a crystal structure is here derived. One first notices that the wavelength of X radiation is comparable with interatomic distances in the crystal and with ion sizes. Therefore, each element of charge $\rho_{\mu}(\mathbf{r})d\mathbf{r}$ in (1.36) works as an emitting source of electromagnetic radiation, with $\rho_{\mu}(\mathbf{r})$ the electric charge density of atom species labeled by μ . Each charge element is located at $\boldsymbol{\rho} = \mathbf{R} + \mathbf{d}_{\mu} + \boldsymbol{\xi}$, where \mathbf{R} is the vector of the direct lattice pointing to a primitive cell, \mathbf{d}_{μ} locates the position of the atomic nucleus of species μ within the cell, and $\boldsymbol{\xi}$ the position of the charge element with respect to the nucleus.

Equation (1.36) can thus be detailed to be:

$$E \propto \sum_{n_1, n_2, n_3} \sum_{\mu} \int d\boldsymbol{\xi} \rho_{\mu}(\boldsymbol{\xi}) e^{i\Delta\mathbf{k} \cdot (\mathbf{R} + \mathbf{d}_{\mu} + \boldsymbol{\xi})} = S_R F_G \quad (1.37a)$$

$$S_R = \sum_{n_1, n_2, n_3} e^{i\Delta\mathbf{k} \cdot \mathbf{R}} \quad (1.37b)$$

$$F_G = \sum_{\mu} f_{\mu}(\Delta\mathbf{k}) e^{i\Delta\mathbf{k} \cdot \mathbf{d}_{\mu}}, \quad (1.37c)$$

with $f(\Delta\mathbf{k}) = \int_{V_a} d\boldsymbol{\xi} \rho_{\mu}(\boldsymbol{\xi}) e^{i\Delta\mathbf{k} \cdot \boldsymbol{\xi}}$ and V_a the atomic or ionic volume of species μ .

In essence, (1.37a)–(1.37c) state that

Concept

The intensity of elastic scattered radiation is composed of the product $S_R F_G$ of two structure factors traceable back to the Bravais lattice and its base. S_R accounts for the arrangement of the Bravais lattice sites. The geometrical factor F_G instead, accounts for the arrangement of the atomic species μ within the base cell, they being identified by an atomic structure factor $f_{\mu}(\Delta\mathbf{k})$.

As already discussed, (1.37a)–(1.37c) neglect the contribution coming from nuclei, since the corresponding scattering radiation has much smaller intensity than that from electrons.

The single terms are now further detailed. Let us start from S_R in (1.37b). Since $\mathbf{R} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3$ and considering that a finite crystal has a limited number of sites, say N_i along the $i = 1, 2, 3$ direction, one obtains:

$$\sum_{n_1, n_2, n_3} e^{i\Delta\mathbf{k} \cdot \mathbf{R}} = \sum_{n_1=0}^{N_1-1} (e^{i\Delta\mathbf{k} \cdot \mathbf{a}_1})^{n_1} \sum_{n_2=0}^{N_2-1} (e^{i\Delta\mathbf{k} \cdot \mathbf{a}_2})^{n_2} \sum_{n_3=0}^{N_3-1} (e^{i\Delta\mathbf{k} \cdot \mathbf{a}_3})^{n_3}. \quad (1.38)$$

Reminding that in geometrical series $\sum_{n=0}^{N-1} x^n = (1 - x^N)/(1 - x)$, one has

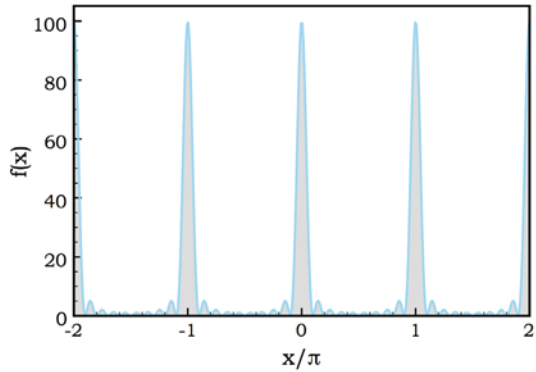
$$\sum_{n_1=0}^{N_1-1} (e^{i\Delta\mathbf{k} \cdot \mathbf{a}_1})^{n_1} = \frac{1 - e^{iN_1 \Delta\mathbf{k} \cdot \mathbf{a}_1}}{1 - e^{i\Delta\mathbf{k} \cdot \mathbf{a}_1}}.$$

The physical quantity measured at the detector is actually the radiation intensity I_f , which is $\propto |E|^2$. Thus,

$$I_f \propto \left| \frac{1 - e^{iN_1 \Delta\mathbf{k} \cdot \mathbf{a}_1}}{1 - e^{i\Delta\mathbf{k} \cdot \mathbf{a}_1}} \right|^2 = \frac{\sin^2(N_1 \Delta\mathbf{k} \cdot \mathbf{a}_1/2)}{\sin^2(\Delta\mathbf{k} \cdot \mathbf{a}_1/2)}. \quad (1.39)$$

Fig. 1.32 displays the overall structure of functions of the type $f(x) = \sin^2(Nx)/\sin^2 x$. This is characterized by principal maxima located at $x = n\pi$, where it assumes the value $f(n\pi) = N^2$, and by secondary maxima shorter than the principal ones. On the other hand, similar results hold for the remaining two sums in (1.38). Thus, the maximum intensity of the scattered radiation occurs whenever $\Delta\mathbf{k} \cdot \mathbf{a}_1 = 2\pi n$,

Fig. 1.32 Elastic scattering of radiation from atoms. The function $f(x) = \sin^2(Nx)/\sin^2(x)$ entering the scattered radiation intensity (1.39) is displayed. For the sake of example, the case with $N = 10$ is plotted. The principal maxima at $x = n\pi$ with n a relative integer number are evident, their height being $f(n\pi) = N^2$



$\Delta \mathbf{k} \cdot \mathbf{a}_2 = 2\pi m$, and $\Delta \mathbf{k} \cdot \mathbf{a}_3 = 2\pi p$. The conditions above are named after Laue. That is, when

$$\Delta \mathbf{k} = \mathbf{G}, \quad (1.40)$$

Concept

The intensity of the scattered radiation is highest whenever $\Delta \mathbf{k}$ fits a vector \mathbf{G} of the reciprocal lattice.

Equation (1.40) results into the consideration that a measure of $\Delta \mathbf{k}$ infers knowledge of the reciprocal-lattice vectors and therefore of the positions of atoms in the direct lattice. In the absence of a base, the intensity of scattered radiation is the same for all the maxima. Thus, a modulation of the intensity along the directions $\Delta \mathbf{k} = \mathbf{G}$ would be a signature that an F_G structure factor plays a role in (1.37a)–(1.37c), inferring information on the configuration of the base and of the atomic structure factors. A few selected examples to illustrate this point are discussed in the following, after stepping the procedure to calculate F_G .

1.6.4 Selected examples of scattering spectra

The connection between theory and experiments is best learned if simple examples are illustrated. To this aim, the case of simple cubic lattices is here analyzed. With reference to (1.37a)–(1.37c), (1.39), and (1.40), the following procedure can be traced:

Procedure

Step 1. Take the conventional cell and identify the positions of whatever n atoms of species labeled by μ . These are represented by the expression $\mathbf{d}_\mu = x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + x_3 \mathbf{a}_3$, with x_1 , x_2 and x_3 being non-integer numbers.

Step 2. For each μ calculate $\mathbf{G} \cdot \mathbf{d}_\mu = 2\pi(r_1x_1 + r_2x_2 + r_3x_3)$, where $\mathbf{G} = r_1\mathbf{b}_1 + r_2\mathbf{b}_2 + r_3\mathbf{b}_3$.

Step 3. Insert the above terms in the sum (1.37c) and calculate it.

Step 4. While calculating the sum, evaluate to exploit possible symmetries.

The above procedure is now ready to be applied to the following examples aimed to determine the scattered intensity for a simple cubic lattice with base.

Examples

Face-centered cubic lattice. Let the cell contain four atoms located at positions $\mathbf{d}_\mu = (x_{1\mu}, x_{2\mu}, x_{3\mu})$ with $\mu = 1, 2, 3$, and 4 given by

$$\mathbf{d}_1 = (0, 0, 0),$$

$$\mathbf{d}_2 = (a/2, a/2, 0),$$

$$\mathbf{d}_3 = (a/2, 0, a/2), \text{ and}$$

$$\mathbf{d}_4 = (0, a/2, a/2),$$

in the conventional cubic cell notation. The geometric factor is

$$F_G = \left| f_1(\mathbf{G}) + f_2(\mathbf{G})e^{i\pi(r_1+r_2)} + f_3(\mathbf{G})e^{i\pi(r_1+r_3)} + f_4(\mathbf{G})e^{i\pi(r_2+r_3)} \right|^2. \quad (1.41)$$

If atoms are all the same along with the corresponding atomic structure factors, these describe a compound crystallizing into a face-centered cubic lattice.

Inspection of (1.41) leads to the conclusion that $F_G \neq 0$ only if r_1, r_2 and r_3 share the same parity.

Examples

Diamond. The above results can be applied to the case of diamond which, as already discussed in Sec. 1.3.1.1, crystallizes into a face-centered cubic lattice with a two-atoms base. The conventional cell of the simple cubic has eight atoms located at $\mathbf{d}_1 \equiv (0, 0, 0)$, $\mathbf{d}_2 \equiv (a/2, a/2, 0)$, $\mathbf{d}_3 \equiv (a/2, 0, a/2)$, $\mathbf{d}_4 \equiv (0, a/2, a/2)$, $\mathbf{d}_5 \equiv (a/4, a/4, a/4)$, $\mathbf{d}_6 \equiv (3a/4, 3a/4, a/4)$, $\mathbf{d}_7 \equiv (3a/4, a/4, 3a/4)$, and $\mathbf{d}_8 \equiv (a/4, 3a/4, 3a/4)$.

One obtains:

$$\begin{aligned} F_G &= |f|^2 \left| 1 + e^{i\pi(r_1+r_2)} + e^{i\pi(r_1+r_3)} + e^{i\pi(r_2+r_3)} + e^{i\frac{\pi}{2}(r_1+r_2+r_3)} \times (1 + e^{i\pi(r_1+r_2)} \right. \\ &\quad \left. + e^{i\pi(r_1+r_3)} + e^{i\pi(r_2+r_3)}) \right|^2 \\ &= |f|^2 \left| 1 + e^{i\pi(r_1+r_2)} + e^{i\pi(r_1+r_3)} + e^{i\pi(r_2+r_3)} \right|^2 \left| 1 + e^{i\frac{\pi}{2}(r_1+r_2+r_3)} \right|^2. \end{aligned} \quad (1.42)$$

This result is consistent with the fact that atoms in diamond are arranged on the sites of two face-centered lattices translated by $(a/4, a/4, a/4)$ with respect to each other. If r_1, r_2 and r_3 were to share the same parity, then

$$\left| 1 + e^{i\frac{\pi}{2}(r_1+r_2+r_3)} \right|^2 = \begin{cases} 4 & \text{if even parity and } r_1 + r_2 + r_3 = 4n \\ 0 & \text{if even parity and } r_1 + r_2 + r_3 = 2n \quad (\text{with } n \text{ odd}) \\ 2 & \text{if odd parity} \end{cases}.$$

Examples

Body-centered cubic lattice. Let the cell contain two atoms located at positions $\mathbf{d}_\mu = (x_{1\mu}, x_{2\mu}, x_{3\mu})$ with $\mu = 1$ and 2. These are given by $\mathbf{d}_1 = (0, 0, 0)$, and $\mathbf{d}_2 = (a/2, a/2, a/2)$, in the conventional cubic cell notation. If atoms are all the same along with the corresponding atomic structure factors, these describe a compound crystallizing into a body-centered cubic lattice. One obtains:

$$F_G = \left| f_1(\mathbf{G}) + f_2(\mathbf{G})e^{-i\pi(r_1+r_2+r_3)} \right|^2. \quad (1.43)$$

Inspection of (1.43) leads to the conclusion that $F_G \neq 0$ only if $r_1 + r_2 + r_3$ is an even integer. For example, metallic sodium crystallizes into a bcc lattice, and does show (200), (110) or (222) reflections, whereas (100) or (111) reflections are absent.

Quick Questions

Q11. The reflection on the (100) plane is absent for the bcc lattice. Provide a physical interpretation.

Answer. For a simple cubic lattice the (100) reflection is observed when the reflections from two consecutive lattice planes separated by a , differ by a 2π phase difference. However, in the bcc lattice, a third plane is interposed between these two, at distance $a/2$. This originates a third reflection, differing by a phase π with respect to the former two: as a result, destructive interference sets in. Notice that similar conditions occur also in CsCl, as in Fig. 1.25. In this case though, the atomic structure factor $f_1 = f_{\text{Cs}}$ is different and of much larger size than the $f_2 = f_{\text{Cl}}$, due to the fact that the two ions contain a significantly different number of electrons.

Scattering spectra and chemical bonds. The example of diamond is one more time amenable to draw a more general consideration. Diamond has a small atomic number Z and four valence electrons. Now, the dependence of the atomic structure factor on \mathbf{G} reflects the distribution of electronic charges of the atom placed inside the crystal. The latter is in general different from the charge distribution of a free atom, since in particular the atomic structure factor takes into account modifications in the distribution of charges of valence electrons. For those materials where the number of valence electrons is small with respect to the total number, then the atomic structure factor of the free atom is a good approximation to be used in calculations also of crystal structures. As seen, this is not the case of diamond. Indeed, if the expression (1.42) were used to interpret the experimental data, a significant discrepancy would show up. The data are instead in good agreement with a model where the formation of the covalent bond between the atom in $(0,0,0)$ and that in $(a/4, a/4, a/4)$ is mimicked by introducing a tiny charge at the point $(a/8, a/8, a/8)$, which would describe the charge displacement towards the intermediate region between the two atoms. A further modulation of the scattered intensity results, as tackled in Problem 1.4. In conclusion,

Concept

Information on the nature of the chemical bonds can in general be inferred from the comparison between experimental data and theoretical expressions for the total scattered intensity.

1.6.5 Bragg's law

An alternative and physically significant interpretation of the angular dependence of the scattered beams from a crystals is due to Bragg, in terms of a diffraction law. Consider a wave impinging on a crystal, that is specular-like reflected by a family of parallel and evenly-spaced planes. Consider the case in which only a small fraction of the wave is reflected by each plane and in which the scattering is just elastic: that is, the reflection does not change the particle wavelength, whatever photon or electron or neutron it be.

Let d be the spacing between the planes within the family, and θ the angle that the \mathbf{k} -vector of the incoming radiation forms with the planes. The difference in the optical path length spanned by two rays reflected by adjacent planes is $2d \sin \theta$. A geometrical sketch of the corresponding situation is illustrated in the left panel of Fig. 1.33. Constructive interference occurs when the optical path difference is a multiple integer of the wavelength. That is, the maximum intensity of the diffracted beam occurs whenever

$$2d \sin \theta = n\lambda . \quad (1.44)$$

This is the statement of Bragg's law. Significantly enough,

Concept

Bragg's law (1.44) is just and precisely a consequence of crystal periodicity: detail about atomic positions and basis composition disappearing whatsoever. Also, Bragg reflection requires wavelengths $\lambda \leq 2d$, that are far away smaller than those of visible light.

Expression (1.44) is fully equivalent to the condition $\Delta \mathbf{k} = \mathbf{G}$ derived in Sec. 1.6.3, which identifies the maxima in the scattered intensity. The proof is straightforward. Indeed, once \mathbf{G} is given, the least-modulus vector \mathbf{G}_0 parallel to \mathbf{G} is given too, namely $\mathbf{G} = h\mathbf{G}_0$, with h an integer. Now, \mathbf{G}_0 uniquely determines the family of lattice planes perpendicular to it, spaced by $d = 2\pi/G_0$. Graphical representation of the vectors as in the right panel of Fig. 1.33, leads to an isosceles triangle with sides \mathbf{k} , \mathbf{k}' and $\mathbf{G} = \mathbf{k}' - \mathbf{k}$. By inspection, one has $G/2 = k \sin \theta = (2\pi\lambda^{-1}) \sin \theta$. This is in turn equivalent to the Bragg's law (1.44), once the correspondence of the integer numbers $n \leftrightarrow h$ is performed.

The next two sections briefly detail the concepts underlying selected experimental techniques which are able to extract the information discussed in Sec. 1.6.2. These include techniques to investigate both crystal structures and nanostructures, the latter requiring special care due to the significance of surface effects.

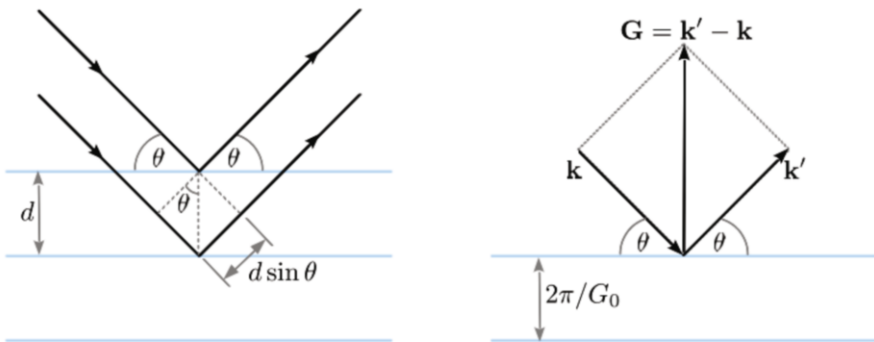


Fig. 1.33 Diffraction from a crystal. Left panel: geometrical illustration of Bragg's law. Right panel: the condition $\Delta \mathbf{k} = \mathbf{G}$ for maximum intensity of the scattered wave is equivalent to Bragg's law (see text)

1.6.6 Determination of the structure of crystals

The determination of the reciprocal lattice vectors is simplified after adopting a procedure that is named Ewald construction. Fig. 1.34 illustrates how it works. Let a lattice point be arbitrarily chosen. The vector \mathbf{k} is first drawn starting from this selected lattice point. The tip of \mathbf{k} is then taken as center, to draw a sphere of radius k . All the lattice points on the sphere surface are those that can be detected after using radiation of wavelength $\lambda = 2\pi/k$. Thus, they could be detected by rotating the detector position along the whole solid angle. The detection of other lattice points would require a tuning of the radiation wavelength. Detector rotation and wavelength tuning are both inconvenient operations in an experimental setup.

This problem is solved by rotating the crystal instead. In this case, the reciprocal lattice vectors that can be detected become those intersecting the Ewald sphere. Use of thin powders containing microcrystals is often equivalent to use a single crystal and rotate it. Microcrystals can indeed be considered as composed of many crystals oriented in all possible directions. By Ewald construction, the reciprocal lattice vectors with size $G = 2k \sin \theta$ compatible with the used wavelength are detected, as shown in Fig. 1.34. As elaborated in Problems 1.5 and 1.12, each lattice has a characteristic ratio of the resulting G -vectors moduli, and thus this procedure is sufficient for a univocal identification of the lattice.

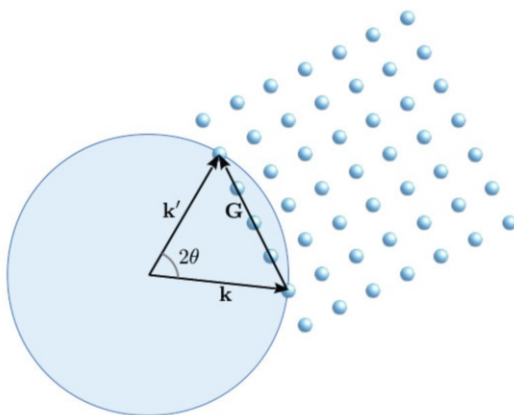


Fig. 1.34 Experimental techniques to determine crystal structures. The Ewald construction is depicted

1.6.7 Determination of the structure of nanosystems

Nanostructures are spatially confined systems, their surface thus playing a significant role. The relevance of surface effects has to be taken care of, while developing techniques suited to determine the structure of nanosystems [7]. This is discussed in the following.

1.6.7.1 X-rays based techniques

One might still consider to use X-rays based techniques. One might think to revert back to expression (1.36) and calculate it for one single nanostructure. However, one single nanostructure would provide a low signal at the detector. Thus, one might want to have many of them in a matrix. At variance with the case of a crystal, here the charges are dispersed in the nanostructure, and therefore the total field at the detector would strongly depend on how the nanostructures are distributed within the matrix.

This complicated problem can be tackled by first considering the simplest situation where the total charge of size Q generating the field to be detected, is schematized by a uniform spherical distribution of radius R_0 . Starting from (1.36), Problem 1.7 elaborates the expression acquired by the scattered field E at the detector. This is

$$E \propto F(qR_0) = \frac{3Q}{qR_0} \left[\frac{\sin(qR_0)}{(qR_0)^2} - \frac{\cos(qR_0)}{qR_0} \right], \quad (1.45)$$

where $q = |\mathbf{k} - \mathbf{k}'| = (4\pi/\lambda)k \sin(\theta/2)$ and θ is the angle between \mathbf{k} and \mathbf{k}' . The detected intensity of the field scattered by the single nanostructure would be $\propto F^2(qR_0)$. For a whole matrix with N nanostructures randomly distributed within it, the variance for the distribution of scattered intensities needs to be calculated. The latter depends in turn on the distribution of electronic charges representing the single nanostructures. To this aim, the variance $(d - d_0)^2$ comes to be helpful, measuring the dispersion of the average charge-density d of the nanostructure with respect to the average charge-density d_0 of the matrix. For a gaussian random distribution of nanostructures, the detected intensity of the scattered field would on average result to be:

$$I(q) = I_0 N (d - d_0)^2 F^2(qR_0), \quad (1.46)$$

as related to the intensity I_0 of the incoming field [8, 9].

Concept

The linear scaling of $I(q)$ with the number of nanostructures N is a general consequence of having a set of incoherent and randomly distributed items.

This is a neat difference with crystals, where the N cells are regularly distributed in the crystal “matrix” and thus coherently contribute to construct the scattered intensity. The latter indeed scales with N^2 , as evidenced in (1.39) and in Fig. 1.32.

An ordered distribution of the nanostructures within the matrix requires a different counting of the contributions from single nanostructures. Instead of calculating the variance of a random distribution of the type (1.45), one has to calculate the sum

$$E \propto F(qR_0) \sum_n e^{i\Delta\mathbf{k} \cdot \mathbf{R}_n} \quad (1.47)$$

over the positions \mathbf{R}_n of the n -th nanostructure inside the matrix. Engineering the position of single nanostructures within a matrix is a tough problem. In those cases

where such an ordered distribution can be realized, the observed diffraction peaks can be investigated by the same procedure as with crystal powders. The advantage here is that one can focus on selected diffraction peaks. Deviations from $F^2(qR_0)$ provide instead information on possible non uniform electronic distributions of nanostructure charges. Since in such a regular matrix of nanostructures the primitive vectors would have a large size, the vectors of the reciprocal lattice would be correspondingly small sized. Thus, the use of matrices of regularly arranged nanostructures provides information on the $q \rightarrow 0$ behavior of physical observables, which would otherwise remain not accessible.

1.6.7.2 Electron spectroscopy

As already discussed, use of X rays in inelastic scattering experiments is not convenient, since the typical incoming photon energies are in the range 10^3 – 10^4 eV, much larger than all other relevant energies into play, thereby leading to excessively small relative energy variation. For inelastic scattering, use of massive particles is much more convenient. These are characterized by de Broglie wavelengths of the order of $\lambda_{dB} \simeq 10^{-8}$ cm. Electron beams with waists of the order of $\simeq 25$ nm² are typically used in this kind of experimental techniques. Here is a schematic illustration of Scanning Electron Microscopy (SEM) and Transmission Electron Microscopy (TEM).

SEM. In Scanning Electron Microscopy electrons with energies > 50 KeV are shot onto the matrix containing the nanostructures. Different phenomena may occur:

- **Elastic scattering** Incoming electrons are back scattered leaving their energy unchanged
- **Emission of secondary electrons.** Electrons from the target atoms are expelled after the interaction with the incoming electrons
- **Emission of bremsstrahlung radiation.** Incoming electrons passing nearby a nucleus are accelerated as skiers who slalom a pole, thereby emitting radiation: this is named after the German term bremsstrahlung for braking radiation
- **Luminescence.** Incoming electrons expel electrons in deeply bound states of target atoms. While returning to its ground state in a typical time of tens of nanoseconds, an X radiation characteristic of the target atom is emitted
- **Production of Auger pairs.** Incoming electrons expel electrons in deeply bound states of target atoms. Electrons in higher states make a transition into these deeper, unoccupied, states, and make a significant amount of energy available. This in turn determines the emission of pairs of electrons, named after Auger

Different mechanisms are characterized by different energy and momentum spectra of outgoing radiation or particles, providing complementary information. For example, the cross section and thus the probability for elastic backscattering increases with the mass atomic number of the target atom, but it does not contain characteristic

signatures of the atomic species. The latter instead show up as peaks in fluorescence or in Auger spectra, on top of the background given by bremsstrahlung radiation.

TEM. In Transmission Electron Microscopy electrons with energies of the order of 100 KeV are shot onto thin targets less than 200 nm thick. Both elastic and inelastic scattering may occur here. In the former case, diffraction peaks are produced. In the case of inelastic scattering, complicated interaction processes take place due to inhomogeneity of the material, presence of boundary interfaces between different grains in the structure, dislocations around which lattice atoms are misaligned, and other defects due to variations of the density of the material. Fig. 1.35 displays a TEM image of a Si nanostructure, along with a representation of the structure model which is compatible with the experimental data. Interestingly enough, the TEM image can distinguish at immediate sight the crystalline part of the structure from the amorphous phase of the matrix, SiO_2 in this specific case.

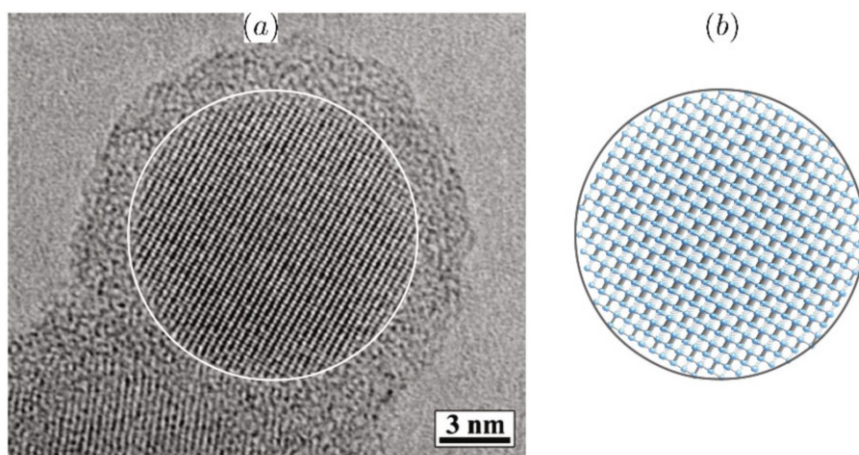


Fig. 1.35 Determination of nanosystem structures by TEM spectroscopy. Example of TEM on a Si nanostructure in a silicon-oxide SiO_2 matrix. (a) Observed TEM image. (b) Model of the structure compatible with the observed image. For graphical reasons the structural model is given with a number of atoms much smaller than the corresponding TEM image in (a) [10]

1.6.7.3 Scanning probe microscopy

More recent techniques are best amenable to investigate surface structures, broadly referred to as Scanning Probe Microscopy (SPM). In typical SPM, a probe tip is mechanically raster-like moved by means of piezoelectric actuators in order to scan the sample surface. Each different type of SPM uses a different concept and observable to probe the surface during the scanning as a function of position. In any event, the resulting observable data are then elaborated, generally providing false-

color images. These techniques can provide both visualization and manipulation of structures at nanometric scales, down to the size of a single atom, namely thousands times below the optical diffraction limit.

Scanning Tunneling Microscope (STM). SPM has begun in 1981 with the invention of the Scanning Tunneling Microscope (STM) and is not surprisingly considered the innovation which has started the nanotechnology era.

Concept

The working concept of STM is based on the tunnel effect: under suited conditions, an electron from an atom on the sample surface has a finite probability of tunneling below the potential barrier between tip and surface, that are generally kept not in contact with each other.

The sensor-tip in an STM thus gauges the current originated by the tunneling of those electrons which are transferred from the surface to the tip or viceversa, after a given voltage is applied between tip and sample in different ways depending on different operational modes. The experimental setup is schematized in Fig. 1.36. A voltage in the range 1–1000 mV is applied between tip and sample, their mutual distance generally being in the interval 0.5–1 nm. Typical tunneling currents are very low, in the order of pA to nA. Appendix 1.9 reminds the calculation of the transmission coefficient T (see 1.97) for an electron across a barrier with height U_0 and thickness d . This is:

$$T = T_0 e^{-2dk_1}, \quad (1.48)$$

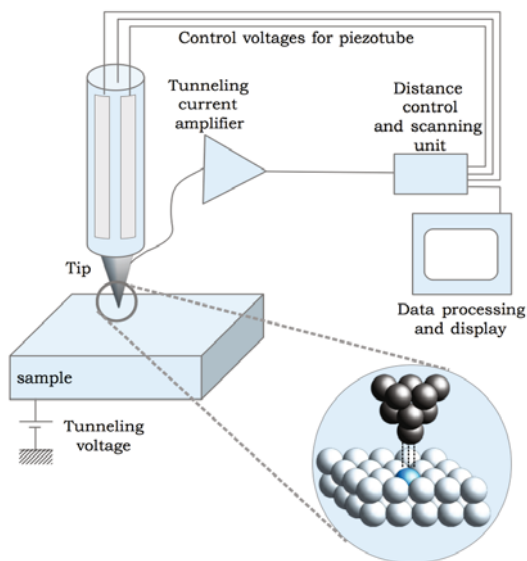


Fig. 1.36 Determination of nanosystem structures by SPM-STM spectroscopy. General scheme of the working setup of a Scanning Tunneling Microscope as described in the main text. The feedback mechanism controlling the tip-sample distance for different operational modes is also shown

where $k_1 = \sqrt{2m(U_0 - E)/\hbar^2}$ is determined by the electron mass m and by the difference between U_0 and the energy E . The validity of (1.48) is limited to the conditions $dk_1 \gg 1$ and $U_0 > E$. The probability current J to find an electron extracted by the sample and moving in the region beyond the barrier is proportional to T and is given by

$$J = \frac{\hbar}{2mi} \left[\psi_R^*(z) \frac{d\Psi_R(z)}{dz} - \Psi_R(z) \frac{d\psi_R^*(z)}{dz} \right], \quad (1.49)$$

in terms of the electron wavefunction $\Psi_R(z)$ near the tip region.

Atomic Force Microscopy (AFM). In other types of SPM the observable can be different. For example,

Concept

In AFM, the working concept is based on the fact that the tip, a tiny lever in fact, is sensitively deflected by the interaction forces between sample and tip, which can be kept either in contact or non-contact.

The sensor-tip in an AFM is actually a cantilever whose displacement is gauged by optical means.

Nanowriting. SPMs can also be exploited to manipulate surfaces, up to positioning on them single atoms or molecules with atomic precision. In this sort of so-called nanoreading and nanowriting, one-dimensional and two-dimensional nanostructures can be engineered.

Typical operations to nanowrite atoms of given species on a surface, are sketched below. First, the substrate surface has to be as free as possible from defects: no impurities, no structural defects. Thus, it has to be cleaned and best taken under vacuum. Cleaning operations can be typically performed by cycles of thermal annealing and sputtering. Annealing uses the concept that a crystal might use energy to re-arrange its own structure. The energy is provided to the crystal under the thermal form, with temperatures as high as 1000 K. Sputtering is a sort of evaporation of surface atoms after shooting them with noble gases. Sputtering is widely used in industrial processes to treat surfaces by vapor depositions. Once the surface has been accurately cleaned, the atoms for nanostructure writing are deposited. Sputtering itself can be exploited to this aim, vaporizing the wanted atoms from a target and then depositing them onto the host material surface. The composition and orientation of the latter are chosen to have a suited crystal structure to adsorb the writing atom: the distance between atoms in the host surface is to be close to that of the writing atoms in their original crystal structure. For example, in the case of nanowriting with Au atoms, the host surface is chosen to be NiAl, a metal with the crystal structure of CsCl. The orientation of NiAl is chosen to be the (100), since on this plane Ni and Al atoms are 0.289 nm apart, closest to the 0.288 nm bond distance of Au atoms in the crystal. This point has been illustrated in [Fig. 1.20](#).

Once the nanowriting atoms have been deposited, they can be manipulated with an SPM tip to write, for example, one chain. The preliminary operation here is to make the SPM tip as small and with known chemical composition as possible, ideally making it to be one single atom. The tip crash technique comes here to hand: the tip is indeed crashed into a selected surface and then retracted, so that it can be covered by well known atoms of the damaged surface. Such tip-crash techniques are for example exploited to dig into or to manufacture a surface, or even to measure its hardness. Pulling-mode operation can thus be used to move and control the atoms in the desired way: when the tip is moved close to the deposited atom, a tip motion pulls as well the atom in the new position. In order for the pulling operation to be effective, a good measurement is required of the distance between tip and surface atom. This is indirectly provided by the tunneling resistance, which after appropriate calibration turns out to be in the order of hundreds of $K\Omega$ for distances of 1–2 Å, and in the range of $M\Omega$ – $G\Omega$ for distances of the order of 10 Å. Operational modes different than pulling can be best suited depending on substrate composition and atom species, and on size and polarity of the applied voltage. For example, in the so-called pushing mode an atom is indeed pushed away from the tip, while in the vertical mode it is first lifted up by the tip and then released into the new position.

1.7 Cohesion in solids

Down to this point concepts and tools have been developed to understand and describe the structure of materials. The problem of the existence of these materials in stable form is now handled. To this aim, a fundamental quantity is the cohesive energy E_c of an atom or molecule in a solid.

Definition

The cohesive energy is the energy needed to take that atom or molecule apart at infinite distance from the solid.

This quantity is related to the total energy E_i . One way to calculate E_c is indeed by difference with the energy E_f of all the system parts as if they were isolated, namely $E_c = E_f - E_i$.

Concept

If $E_c = E_f - E_i > 0$, the energy is larger when the parts are separated from each other than when they are stuck together, thus the solid is stable. If $E_c < 0$ instead, the system spontaneously tends to break up.

An alternative way to calculate the cohesive energy is to identify it as the total energy per particle $E_c = -E_i/N$, where the minus sign is introduced by convention

since E_i is negative. While using this second way, one has to consider that E_i does not necessarily scale with N , and thus the ratio E_i/N might maintain a dependence on N . The two ways produce different results. Between the two, the former is better appropriate, the latter is simpler.

Typical values of cohesive energies range in a wide interval, from 0.1 eV in molecular materials to 1 eV for metals, up to 10 eV for ionic or covalent crystals. These values can be significantly different in nanostructures, due to finite size effects and to possible modifications of the chemical bonds nature. Evidently, the order of magnitude of the cohesive energy reflects the nature and strength of the interactions which are responsible for crystal formation. The melting temperature is often a good indicator for the relative bond strength of different materials.

This section deals with selected interaction mechanisms holding atoms and molecules together to form either a crystal or a nanostructure. The mechanisms are discussed in order of increasing cohesive energies they lead to. In fact, each of the described mechanisms rarely acts alone: it can rather be predominant, while cooperatively contributing along with others. In the following, each of them is separately discussed to help the comprehension [6]. Throughout the treatment, the interactions generally contribute to the total energy in a much larger extent than dynamical effects, and thus all the kinetic-energy terms are consistently neglected.

1.7.1 Phenomenology

Bonds in solids or molecules are described by an interaction potential.

Concept

The interaction potential is usually obtained as a superposition of attractive and repulsive contributions.

The repulsive part accounts for the impenetrability of matter. When two atoms get closer to each other, their charge distributions tend to overlap. In other words, electrons from the first atom more likely occupy states belonging to the second atom and viceversa. Pauli exclusion principle though, prevents occupation of the already occupied lowest energy states: thus, occupation of higher energy, excited states, is favored. The resulting charge distribution overlap increases the total energy of the system and gives rise to a repulsive contribution in the interaction potential.

The nature and origin of the attractive part varies. Usually, bonds can be schematically classified as follows. (i) Ionic bonds are originated by electron transfer from an electropositive to an electronegative atom: the resulting positive and negative ions feeling a Coulomb attraction potential. (ii) Covalent bonds are originated by one or more electrons sharing between first neighbor atoms. (iii) Metallic bonds are originated by a sort of valence-electrons removal from the host atoms: as a result, electrons are delocalized within the material, where they balance the charge of the

underlying positive ion-core background. (iv) Van der Waals bonds are originated by quantum fluctuations of the microscopic dipole moments of atoms or molecules.

Quick Questions

Q12. In a virtual experiment, two H atoms are approached to each other.

In the limit of full overlap, the ground state is that of He atom with $1s^2$ configuration, in which the two electrons have opposite spins, and total energy ~ -78.9 eV. However, if the two electrons are forced to have the same spin, an excited state is obtained with electronic configuration $1s^1 2s^1$, whose total energy is ~ -59.38 eV. Explain the origin of the energy difference.

Answer. The difference between the two energies can be traced back to a repulsive effect due to the Pauli exclusion principle.

1.7.2 Van der Waals, dipolar, and hydrogen bonds

On phenomenological grounds, atoms or molecules forming a solid where Van der Waals, dipolar, or hydrogen-like bonds are predominant, do not alter their structure when in the crystal.

Concept

Van der Waals mechanisms originate from the interactions between dipoles which can be induced on atoms or molecules.

Within a classical approach, a molecule with electric dipole moment p is a source of electric field $E \propto p/r^3$. This field generates in turn an induced dipole $p_{in} \propto \alpha E \propto \alpha p/r^3$ on a second molecule at distance r , where α is its polarizability. The two dipoles thus interact with each other, each being a dipole in the electric field generated by the other. Due to the intrinsic mechanism of induced polarization, the interaction is attractive. The corresponding energy is $-p_{in}p/r^3 \propto -\alpha p^2/r^6$, varying as $1/r^6$. At shorter distances between molecules, a repulsive interaction becomes predominant instead, due to repulsive correlations between electrons in the molecules and to Pauli exclusion principle.

Atoms or molecules crystallizing with Van der Waals-like bonds are for example the noble gases, and also molecules like O_2 , N_2 , F_2 , ..., methane-like compounds such as CH_4 , CCl_4 and the like, more complex structures of the type $CH_3(CH_2)_nCH_3$, and selected polymers. In the different cases, the Van der Waals bond acts between the molecules, whereas the bonds keeping together the single molecule might have different nature: covalent in O_2 , N_2 , F_2 and the like, polar in CCl_4 , CF_4 , CBr_4 and the like. In the case of more complex compounds, additional complexity is introduced by different geometrical arrangements, running from linear

chains to packed structures in spherical shape. Van der Waals bonds occur as well in planar structures like graphite, V_2O_5 , MoO_3 , MoS_2 , ..., acting predominantly in the direction perpendicular to the planes.

The melting temperature is an useful indicator of cohesion. All noble gases are characterized by increasing melting temperatures with increasing cohesive energy: He represents an exception, since at very low temperatures it solidifies only under pressure. In crystals like $CH_3(CH_2)_nCH_3$, the melting temperature is an increasing function of the molecular weight of element $(CH_2)_n$, and depends on the geometrical form of the material.

The understanding of Van der Waals interactions within a quantum approach is now being considered, besides the above classical viewpoint. In the following, a justification of the $1/r^6$ behavior is discussed, after starting from a simple situation, where two atoms A and B, say of noble gases, are brought close to each other. The Hamiltonian for this system of two atoms is

$$H = H_0(A) + H_0(B) + H_{AB} , \quad (1.50)$$

where $H_0(A)$ and $H_0(B)$ are the Hamiltonians for atoms A and B if they were isolated, and H_{AB} represents the interaction between them. In more details, one has

$$\begin{aligned} H_0(A) &= \sum_{i=1}^N \left(\frac{p_i^2}{2m} - \frac{Ze^2}{|\mathbf{r}_{Ai} - \mathbf{R}_A|} \right) + \frac{1}{2} \sum_{\substack{i,i'=1, \\ i \neq i'}}^N \frac{e^2}{|\mathbf{r}_{Ai} - \mathbf{r}_{Ai'}|} , \\ H_0(B) &= \sum_{j=1}^N \left(\frac{p_j^2}{2m} - \frac{Ze^2}{|\mathbf{r}_{Bj} - \mathbf{R}_B|} \right) + \frac{1}{2} \sum_{\substack{j,j'=1, \\ j \neq j'}}^N \frac{e^2}{|\mathbf{r}_{Bj} - \mathbf{r}_{Bj'}|} \\ H_{AB} &= \sum_{i,j=1}^N \frac{e^2}{|\mathbf{r}_{Ai} - \mathbf{r}_{Bj}|} + \frac{Z^2 e^2}{|\mathbf{R}_A - \mathbf{R}_B|} - \sum_{j=1}^N \frac{Ze^2}{|\mathbf{r}_{Bj} - \mathbf{R}_A|} - \sum_{i=1}^N \frac{Ze^2}{|\mathbf{r}_{Ai} - \mathbf{R}_B|} . \end{aligned} \quad (1.51)$$

That is, $H_0(A)$ ($H_0(B)$) is composed of the kinetic energies of the N electrons with mass m , by the Coulomb interaction between each electron at position \mathbf{r}_{Ai} (\mathbf{r}_{Bi}) with the corresponding nucleus with atomic number Z at position \mathbf{R}_A (\mathbf{R}_B), and by the Coulomb interaction between each pair of electrons within the same atom. Notice the insertion of the factor of $1/2$ in front of the $\sum_{i,i'}$ and of the $\sum_{j,j'}$, to avoid double

counting. The interaction term H_{AB} accounts instead for all the cross terms, that are the interaction between A and B electrons, A and B nuclei, and between electrons of atom A with nucleus B and viceversa. Under the conditions $|\mathbf{R}_A - \mathbf{R}_B| \gg r_{Ai}$, r_{Bj} , where the two atoms A and B are separated by a distance long on the scale of all the other relevant lengths, perturbation theory can be applied. In particular, the Hamiltonian $H_0(A) + H_0(B)$, corresponding to the two atoms as if they were isolated, can be considered as the unperturbed part, on which the perturbation H_{AB} acts. The ground state of the unperturbed system is represented by the product of the two ground-state wavefunctions $\phi_0(A)$ and $\phi_0(B)$ of atoms in \mathbf{R}_A and \mathbf{R}_B as if they were isolated at very long distance from each other. It is thus convenient to expand

H_{AB} with respect to the variables $\boldsymbol{\eta}_{Bj} = \mathbf{r}_{Bj} - \mathbf{R}_B$ and $\boldsymbol{\xi}_{Ai} = \mathbf{r}_{Ai} - \mathbf{R}_A$, assuming small values $|\boldsymbol{\xi}_{Ai}|/|\mathbf{R}_A - \mathbf{R}_B| \ll 1$ and $|\boldsymbol{\eta}_{Bj}|/|\mathbf{R}_A - \mathbf{R}_B| \ll 1$. To second order in $\boldsymbol{\xi}_{Ai}$ and $\boldsymbol{\eta}_{Bj}$, Problem 1.8 shows that:

$$H' = -e^2 \sum_{\alpha, \beta=1}^3 \frac{\partial^2}{\partial R_{A,\alpha} \partial R_{A,\beta}} \frac{1}{|\mathbf{R}_A - \mathbf{R}_B|} \left(\sum_{j=1}^N \eta_{Bj,\alpha} \right) \left(\sum_{i=1}^N \xi_{Ai,\beta} \right), \quad (1.52)$$

where indexes α and β refer to the three vector components, and the choice has been taken to derive $1/|\mathbf{R}_A - \mathbf{R}_B|$ with respect to \mathbf{R}_A (the choice is immaterial).

The form (1.52) acquired by H_{AB} has a first relevant consequence: to first perturbative order, the matrix element $\langle \phi_0(i) \phi_0(j) | H' | \phi_0(i) \phi_0(j) \rangle = 0$, since $|\phi_0(A) \phi_0(B)|^2$ and $\left(\sum_{j=1}^N \eta_{Bj,\alpha} \right) \left(\sum_{i=1}^N \xi_{Ai,\beta} \right)$ have opposite parity with respect to space symmetry, the former being even and the latter odd. To second order instead, suited intermediate states exist, which make the matrix elements non-vanishing. In this case, calculated perturbative term is seen to contribute to the ground state with a negative sign, and to depend on the distance $|\mathbf{R}_A - \mathbf{R}_B|$ between the nuclei according to the $|\mathbf{R}_A - \mathbf{R}_B|^{-6}$ law.

In any event, whatever classical or quantum microscopic description one is considering, an useful analytical representation of the full Van der Waals interaction potential for noble gases is derived. This is named after Lennard and Jones:

Concept

The two-parameter potential

$$U(R) = \varepsilon \left[\left(\frac{\sigma}{R} \right)^{12} - 2 \left(\frac{\sigma}{R} \right)^6 \right]. \quad (1.53)$$

with σ and ε specific of the given species, represents a simple and useful expression amenable to joint theoretical and experimental investigation.

In (1.53) the first term accounts for the repulsion between electrons in different atoms, and the second for the attractive term discussed so far. Overall, $U(R)$ in (1.53) has a short-range repulsive term modeled $\propto 1/R^{12}$, due to the interaction between electrons in different atoms, the long-range attractive tail $\propto 1/R^6$ discussed so far, and an attractive well at intermediate range. The physical meaning of the parameters σ and ε is immediate after studying this function and looking for the value R_0 at which the minimum of the well is located, that is calculating $\partial U(R)/(\partial R) = 0$. It is easily found that $\sigma = R_0$, representing the distance R_0 at which mechanical equilibrium occurs, the value of the interaction energy at this point R_0 being $U(R_0) = -\varepsilon$. Table 1.7 reports the values of ε and σ for selected noble gases, whereas Fig. 1.37 displays a plot of $U(R)$ along with its decomposition in long-range attractive and short-range repulsive parts.

Expression (1.53) easily leads to calculate the total energy U_i of the i -th atom in the field of all the others in the crystal. Neglecting the kinetic energy as stated at the

Table 1.7 Cohesion in Van der Waals crystals. The Lennard-Jones potential: values of ϵ and σ are listed for the selected noble gases in first column (from [3])

| Element | ϵ (eV) | σ (Å) |
|---------|-----------------|--------------|
| Ne | 0.0024 | 3.25 |
| Ar | 0.0098 | 3.87 |
| Kr | 0.0135 | 4.11 |
| Xe | 0.0185 | 4.46 |

beginning, this is:

$$U_i = \epsilon \left[\sum_{j \neq i} \left(\frac{\sigma}{p_{ij}R} \right)^{12} - 2 \sum_{j \neq i} \left(\frac{\sigma}{p_{ij}R} \right)^6 \right],$$

where R is now the first-neighbors distance and $p_{ij}R$ the distance of the remaining atoms from the given i -th. Within this approximation scheme, the total energy is obtained after summing the U_i over all the N atoms. This is:

$$U(R) = \frac{1}{2}N\epsilon \left[\sum_{j \neq i} \left(\frac{\sigma}{p_{ij}R} \right)^{12} - 2 \sum_{j \neq i} \left(\frac{\sigma}{p_{ij}R} \right)^6 \right]. \tag{1.54}$$

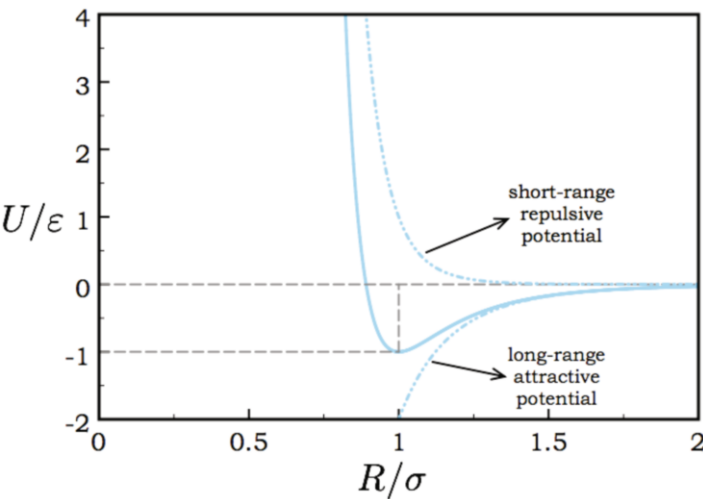


Fig. 1.37 Bonding in solids. A sketch of the Lennard-Jones potential (solid line), resulting from the sum of the attractive long-range and the repulsive short-range components (dashed lines, as in the legend). Notice the quite rapid variation of the curve at distances $R < \sigma$

Note that expression (1.54) linearly scales with the number of atoms N . The advantage of such a writing is that it is suited to all crystal structures with the Van der Waals interactions, the details of the structure being embodied by the numbers p_{ij} .

As to noble gases for example, they crystallize into a face-centered cubic lattice. Thus:

$$U(R) = \frac{1}{2}N\epsilon \left[A_{12} \left(\frac{\sigma}{R} \right)^{12} - 2A_6 \left(\frac{\sigma}{R} \right)^6 \right], \quad (1.55)$$

with

$$\begin{aligned} A_{12} &= \sum_{j \neq i} \left(\frac{1}{p_{ij}} \right)^{12} = 12 + \frac{6}{(\sqrt{2})^{12}} + \frac{24}{(\sqrt{3})^{12}} + \dots = 12.13 \\ A_6 &= \sum_{j \neq i} \left(\frac{1}{p_{ij}} \right)^6 = 12 + \frac{6}{(\sqrt{2})^6} + \frac{24}{(\sqrt{3})^6} + \dots = 14.45. \end{aligned} \quad (1.56)$$

The equilibrium distance between first neighbors disagrees with the values of σ and ϵ as a consequence of the p_{ij} factors. Indeed the condition $\partial U(R)/(\partial R) = 0$ leads to

$$R_0 = \left(\frac{A_{12}}{A_6} \right)^{\frac{1}{6}} \sigma = 0.971 \sigma, \quad (1.57)$$

and to the interaction energy at equilibrium

$$U(R_0) = -\frac{1}{2}N\epsilon \frac{A_6^2}{A_{12}} = -8.61N\epsilon. \quad (1.58)$$

Thus, $U(R_0)$ embodies lots of information: a linear scaling with N , the parameter ϵ identifying the Lennard-Jones potential, and a numerical pre-factor characteristic of the crystal structure. The cohesive energy resulting from the energy per particle is $E_c = -U(R_0)/N = 8.61\epsilon$. This is not the only useful information. Indeed,

Concept

The knowledge of $U(R)$ provides information on cohesive energy. It also provides information on mechanical system properties, precisely its response to a squeezing. This is given by the compressibility coefficient B_0 of the material.

At zero temperature the compressibility is determined by the relative rapidity with which the pressure changes as a consequence of a volume change. The pressure is in turn a measure of how much rapidly the total energy changes after a change in volume, that is $P = -(\partial U / \partial V)|_T$. All together, the compressibility is:

$$B_0 = -V \left(\frac{\partial P}{\partial V} \right)_T = V \left(\frac{\partial^2 U}{\partial V^2} \right)_T. \quad (1.59)$$

Since a face-centered lattice has volume $V = R^3 N / \sqrt{2}$ in terms of the first-neighbors distance R , one has:

$$B_0 = 4\sqrt{2} \frac{\varepsilon}{\sigma^3} A_6 \left(\frac{A_6}{A_{12}} \right)^{\frac{3}{2}} = 106.3 \frac{\varepsilon}{\sigma^3} . \tag{1.60}$$

The compressibility depends on both ε and σ parameters determining the Lennard-Jones potential. Table 1.8 lists first-neighbors distance, cohesive energy per atom and compressibility coefficient for selected crystals of noble gases, along with the comparison with the corresponding experimental data.

The calculation of the cohesive energy and related quantities is to be modified whenever additional interactions, different from Van der Waals, are significant. In crystals of CH₃Cl-like molecules, the latter have a non-vanishing permanent dipole, which introduces an interaction different from that due to induced dipoles. In crystals characterized also by significant hydrogen-like bonds, the energy required to ionize the H atom is balanced out by the electrostatic interactions of the hydrogen proton with the same other atoms (molecules). All these kind of effects have to be taken into account while calculating the cohesive energy.

Quick Questions

Q13. Which is, if any, a notable difference between Van der Waals interactions in crystals of noble gases and in crystals of CH₃Cl-like molecules?
Answer. While the nature of Van der Waals forces is always related to dipole-dipole interactions, in the former case the dipoles are induced, in the latter are permanent.

As a rule of thumb, the van der Waals bond strength is an increasing function of the number of electrons: the larger the number of electrons, the larger is the modulus of the induced dipole. In the case of molecular solids, the bond strength also depends on the molecule aspect ratio: the larger is the molecule aspect ratio, the larger is the number of electrons in a given molecule close to the electrons of an adjacent molecule. Such trends are confirmed by a number of experiments able to

Table 1.8 Cohesion in Van der Waals crystals (from [3]). First-neighbors distance R_0 , cohesive energy E_c and compressibility coefficient B_0 for the selected crystals of noble gases in the first column. For each observable in columns 2–4, the first number represents the experimental data and the second one the theoretical value [11, 12, 13]

| Element | R_0 (Å) | E_c (eV/atom) | B_0 (10^{-10} dine/cm ²) |
|---------|-------------|-----------------|---|
| Ne | 3.13 — 2.99 | 0.02 — 0.027 | 1.1 — 1.81 |
| Ar | 3.76 — 3.71 | 0.08 — 0.089 | 2.7 — 3.18 |
| Kr | 4.01 — 3.98 | 0.116 — 0.120 | 3.5 — 3.46 |
| Xe | 4.35 — 4.34 | 0.17 — 0.172 | 3.6 — 3.81 |

point out the dependence of the melting point on atomic number, molecular weight, and the like.

1.7.3 Ionic bonds

On phenomenological grounds, ionic crystals are characterized by reduced electric conductivity at medium-to-low temperatures, large ionic conductivity at high temperature, along with large infrared absorption capability. Ionic bonds occur for example in crystals with an alkali metal, say Na, and an halogen atom, say Cl, in the cell. Schematically, the alkali metal gives up its only valence electron with a very low ionization energy, which is instead picked up by the halogen atom, which has strong electronegativity with related electron affinity. The balance between the ionization energy and the electron affinity is provided by the electrostatic interaction energy between the positive and negative pair of ions. In NaCl for example, the Na ionization energy is about 5.14 eV and the Cl electronic affinity about 3.71 eV. Therefore, the energy required to form the two isolated ions is $5.14 - 3.71 \text{ eV} = 1.43 \text{ eV}$ and the ion formation process is not favored. On the other hand, the interaction energy between the two ions at the equilibrium distance of about $2.8 \cdot 10^8 \text{ cm}$ is about 5 eV. In this picture, the distance between the ions should shorten with increasing the binding energy. Observations evidence that this is not the case, since a too close approach of the two atoms is prevented by both a repulsive electrostatic interactions between their core electrons, and by the Pauli exclusion principle. Other effects need as well to be accounted for, like the dependence of the number of first neighbors from the size of the largest ion. To conclude these phenomenological considerations, the classical picture given above is not fully adequate to explain the crystal stability.

Concept

The largest part of the cohesive energy in an ionic crystal is composed of the electrostatic potential energy of the single ion generated by all the others, both of like and unlike charge.

Whatever classical or quantum microscopic description one is considering, the calculation of the electrostatic energy is not so easy. The final result is here given, it being expressed in the very simple form

$$E_p = -\frac{\alpha_m e^2}{R} \quad (1.61)$$

given in terms of the first-neighbors distance R and the electron charge e , assuming that the atoms be ionized for only one missing or additional electron. The energy strength parameter α_m is the Madelung constant, whose specific value embodies

characteristics of the crystal structure. A straightforward procedure to determine α_m is as follows:

Procedure

Step 1. Arbitrarily select an ion in the cell.

Step 2. Consider a cube with side nR surrounding the selected ion, with $n \geq 4$ a sufficiently large integer number.

Step 3. Write down the contributions to the electrostatic potential from each single ion using the following recipe, according to the ion position with respect to the chosen cube. The contribution gets a prefactor: 1 if the ion is inside the cube, 0 if it is outside, $1/2$ if it is on a face, $1/4$ if on a side, and $1/8$ if on a vertex. These prefactors are originated by simple geometrical inspection of the ions arrangement on the cube: a face would cut the ion in two, a side in four, and a vertex in eight parts.

Step 4. Sum all the contribution over, to obtain the electrostatic energy.

Examples

Consider a one-dimensional chain of ions with period a and two ions per cell located at 0 and $d \leq \frac{a}{2}$. The calculation of the Madelung constant can be carried out as follows. The positive and negative ions are to be first identified. The positive are at positions $x_n = na$, with $-\infty < n < \infty$, and the negative ones at $y_n = na + d$. Note that the origin is occupied by a positive ion if $d \leq \frac{a}{2}$. Otherwise, one can locate a negative ion at the origin, so that in any case the first neighbors be at $d \leq \frac{a}{2}$. Second, the electrostatic potential V at point $x = 0$ has to be calculated. This is:

$$\begin{aligned}
 V &= -\frac{e}{d} + e \sum_{n=1}^{\infty} \left(\frac{1}{na} - \frac{1}{na+d} \right) + e \sum_{n=-\infty}^{-1} \left(\frac{1}{|na|} - \frac{1}{|na+d|} \right) \\
 &= -\frac{e}{d} + e \sum_{n=1}^{\infty} \left(\frac{1}{na} - \frac{1}{na+d} \right) + e \sum_{n=1}^{\infty} \left(\frac{1}{na} - \frac{1}{na-d} \right) \\
 &= -\frac{e}{d} + ed \sum_{n=1}^{\infty} \frac{1}{na} \left(\frac{1}{na+d} - \frac{1}{na-d} \right) \\
 &= -\frac{e}{d} - 2ed^2 \sum_{n=1}^{\infty} \frac{1}{na} \frac{1}{(na)^2 - d^2} .
 \end{aligned} \tag{1.62}$$

Given $\sigma = a/d \geq 2$, the Madelung constant turns out to be

$$\alpha_m = +1 + 2 \sum_{n=1}^{\infty} \frac{1}{n\sigma} \frac{1}{(n\sigma)^2 - 1} .$$

Since

$$\begin{aligned} \frac{1}{n\sigma} \frac{1}{(n\sigma)^2 - 1} &= \frac{1}{n^3\sigma^3} + \left(\frac{1}{n\sigma} \frac{1}{(n\sigma)^2 - 1} - \frac{1}{n^3\sigma^3} \right) = \frac{1}{n^3\sigma^3} + \frac{1}{n^3\sigma^3 [(n\sigma)^2 - 1]} \\ &= \frac{1}{n^3\sigma^3} + \frac{1}{n^5\sigma^5} + \left(\frac{1}{n^3\sigma^3 [(n\sigma)^2 - 1]} - \frac{1}{n^5\sigma^5} \right), \text{etc.} \end{aligned} \quad (1.63)$$

one finally finds that

$$\alpha_m = +1 + 2 \sum_{p=1}^{\infty} \frac{1}{\sigma^{2p+1}} \sum_{n=1}^{\infty} \frac{1}{n^{2p+1}}. \quad (1.64)$$

From (1.64), α_m can be determined at the desired level of approximation. A few example values are listed in [Table 1.9](#).

Table 1.9 Madelung constant for a 1D chain of ions at different level of approximation

| σ | p | α_m |
|----------|----------|------------|
| 2 | ≤ 1 | 1.3005 |
| 2 | ≤ 2 | 1.3653 |
| 2 | ≤ 3 | 1.3810 |
| 2 | ≤ 4 | 1.3849 |
| 2 | ≤ 5 | 1.3859 |
| 2 | ≤ 6 | 1.3861 |
| 2 | ≤ 7 | 1.3862 |

Similar calculations for cubic systems, such as body-centered (bcc), face-centered (fcc) and zincblend or diamond-like structures yield:

$$\alpha_m = \begin{cases} 1.7627 & \text{for bcc structures, CsCl-like} \\ 1.7476 & \text{for fcc structures, NaCl-like} \\ 1.6381 & \text{for ZnS or diamond-like structures.} \end{cases} \quad (1.65)$$

In fact, note that the values of α_m have no significant variations within the three structures.

Once its main attractive contributing term has been evaluated for the single ion as in (1.61), the total energy can in turn be calculated after assuming to have N positive and N negative ions, each in the field of all the others. In addition, a term accounting for the short-range repulsive interactions is to be introduced. Within this phenomenological approach and the above approximations, the total energy once more scales with the number N of ions and can be expressed as

Concept

$$U(R) = N \left(\frac{\lambda}{R^n} - \frac{\alpha_m e^2}{R} \right), \quad (1.66)$$

where the first term accounts for the short-range repulsion and the second for the Madelung electrostatic attraction.

Once again, expression (1.66) neglects all the kinetic-energy contributions.

The equilibrium distance between the ions is again determined from the condition $(\partial U(R)/\partial R)_{R_0} = 0$, yielding the expression for λ

$$\lambda = \frac{\alpha_m e^2}{n} R_0^{n-1}, \quad (1.67)$$

and then the cohesive energy E_c as the total energy per particle

$$E_c = -\frac{U(R_0)}{2N} = \frac{\alpha_m e^2}{2R_0} \left(1 - \frac{1}{n} \right). \quad (1.68)$$

Since $n \simeq 10$, the result (1.68) confirms the initial physically sound statement, according to which the largest contribution to cohesion comes from the electrostatic energy. This is much better the case for crystals with increasingly larger Madelung constants.

Once the total energy is given, the compressibility can be calculated from (1.59), in much the same way that is been used for Van der Waals compounds. For example, a crystal of the NaCl type with $V = 2NR^3$ would have the compressibility

$$B_0 = (n-1) \frac{e^2 \alpha_m}{18R_0^4}. \quad (1.69)$$

Table 1.10 lists the values of the cohesive energy E_c and of the equilibrium distance R_0 for CsCl and NaCl ionic crystals, that are largely representative for this type of ionic crystals.

Table 1.10 Cohesion in ionic crystals (from [3]). Cohesive energy E_c per ion and equilibrium distance R_0 for CsCl and NaCl. In the second column, the first number represents the experimental and the second one the theoretical value: the latter is calculated from (1.68) as $2E_c$

| Element | E_c (eV per pair) | R_0 (Å) |
|---------|---------------------|-----------|
| CsCl | 6.40—6.83 | 3.571 |
| NaCl | 8.03—8.14 | 2.820 |

In this case too, additional effects could be simultaneously relevant, besides ionic bonding. Two materials, LiCl and SrO, deserve a special mention as a notable example of this situation. LiCl is composed of elements from the I-VII groups, while SrO from II-VI. Both share the property of being polar, with first-neighbors distances $\simeq 2.570$ and 2.580 Å, and cohesive energies $E_c \simeq 8.84$ and 33.34 eV/atom, respectively. The two crystals do not share the same structure. As already learned, the Madelung constants are not significantly different, however the cohesive energy of SrO is about four times that of LiCl. From the above observations one may conclude that the difference in cohesive energy is largely due to the polar character of the bond in SrO.

Quick Questions

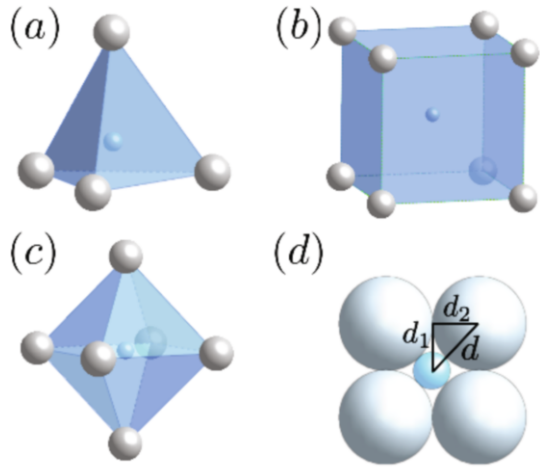
Q14. Consider an ionic compound in which the squared charge of the valence electrons is effectively σe^2 . How λ has to be changed so that R_0 be the same? Which is the scaling behavior of the cohesive energy? If the cohesion energy does not change, how R_0 and λ should scale?

Answer: When $e^2 \rightarrow \sigma e^2$ and R_0 is not changed, it results $\lambda \rightarrow \sigma \lambda$ and $E \rightarrow \sigma E$. On the other hand, if the energy is not changed, $R_0 \rightarrow \sigma R_0$ and $\lambda \rightarrow \sigma^n \lambda$.

Ionic bonds and crystal structure. As a last remark, because of the isotropic nature of Coulomb interactions which the ionic bond relies on, ionic solids would be expected to crystallize in closed packed structures, which maximize the number of first neighbors, or else the number of interactions. Nonetheless, the coordination number in ionic solids is found to depend on the ratio of the anion-to-cation sizes. Defining R_1 (R_2) as the cation (anion) radius, steric effects lead to the formation of different structures depending on the ratio $\delta = R_1/R_2$, as discussed in Sec. 1.3.2. Given the coordination number, a minimum value δ_{\min} exists, below which the structure becomes unstable, because repulsive forces predominate over the attractive ones. In a hard sphere model, such a minimum corresponds to the situation in which the cation is in contact with all its neighbour anions and the anions are in contact with one other. Common configurations are: 3-fold with $\delta_{\min}=0.15$, 4-fold with $\delta_{\min}=0.22$, 6-fold with $\delta_{\min}=0.41$, and 8-fold with $\delta_{\min}=0.73$. These correspond to trigonal, tetrahedral, octahedral and cubic geometries, respectively, as shown in Fig. 1.38.

Consider for example an ionic compound with octahedral coordination, as depicted in Fig. 1.38 (c). The calculation of δ_{\min} is performed in the limiting case in which all the neighbor anions contact one another and the cation is in contact with all its neighbour anions, see Fig. 1.38 (d). One obtains: $d_1 = d_2 = R_2$, $d = \sqrt{d_1^2 + d_2^2} = \sqrt{2}R_2$. Since $d = R_1 + R_2$, we finally get: $R_1 + R_2 = \sqrt{2}R_2 \Rightarrow \delta_{\min} = R_1/R_2 = \sqrt{2} - 1 = 0.414$.

Fig. 1.38 Sketch of common geometries found in ionic compounds: (a) tetrahedral, (b) cubic, and (c) octahedral. In (d) the geometric construction needed to calculate δ_{\min} for the octahedral coordination is shown (see text). Large gray spheres represent the anions and small cyan spheres the cations



1.7.4 Covalent bonds

On phenomenological grounds, covalent crystals can be characterized by a complexity higher than Van der Waals or ionic crystals, as a result of their bonding nature. Sometimes one can determine the cohesive energy in two steps, first calculating the energy of the single molecule and eventually that originated by the interaction between the molecules in the crystal, in principle different from the former. Sometimes however, the two interactions are to be treated simultaneously on the same footing.

Concept

A covalent bond in a, say biatomic, molecule results from an electronic configuration where the wavefunctions of one or two electrons have amplitude probabilities delocalized all over the molecule. The kinetic energy associated to the two-electron orbital is reduced as a consequence of delocalization, an effect which can be traced back to the Heisenberg principle. So does the potential energy, since the two electrons feel the attraction of many atoms.

In the following, the simpler case is treated, where the crystal is composed of biatomic covalent molecules, interacting with each other via either Van der Waals or ionic potentials.

In a quantum approach, a molecular covalent bonding between a pair of atoms 1 and 2 is described by diagonalizing the Hamiltonian H describing the two atoms, in the subspace composed of the orthogonal wavefunctions ψ_1 and ψ_2 of the two isolated atoms, with $\langle \psi_1 | \psi_2 \rangle = 0$. The molecule orbital is a linear combination of the two, that is

$$\psi = u_1 \psi_1 + u_2 \psi_2 . \quad (1.70)$$

Diagonalization of the Hamiltonian proceeds by solving the equation $H\psi = \varepsilon\psi$ to find eigenvectors u_1 and u_2 and eigenvalues ε . To this aim, one forms $\langle \psi | H | \psi \rangle$, leading to the linear system of equations:

$$\begin{aligned} (H_{11} - \varepsilon)u_1 + H_{12}u_2 &= 0 \\ H_{21}u_1 + (H_{22} - \varepsilon)u_2 &= 0 \end{aligned} \quad (1.71)$$

with

$$H_{ij} = \langle \psi_i | H | \psi_j \rangle = \int d\mathbf{r} \psi_i^*(\mathbf{r}) H \psi_j(\mathbf{r}) \quad i, j = 1, 2.$$

For the sake of simplicity, the off-diagonal matrix elements $H_{12} = H_{21}^* = V_2$ can be assumed to be real. The following cases are in order.

1. $H_{11} = H_{22} = \varepsilon_0$. The eigenvalues ε_{\pm} are

$$\varepsilon_{\pm} = \varepsilon_0 \pm |V_2|, \quad (1.72)$$

and the coefficients u_1 and u_2

$$\frac{u_1}{u_2} = -\frac{V_2}{\varepsilon_0 - \varepsilon_{\pm}} = \frac{V_2}{\pm |V_2|}. \quad (1.73)$$

The lowest energy level corresponds to the negative sign, the highest energy level to the positive one. Two sub-cases follow:

- If $V_2 < 0$, the lowest energy state has $u_1/u_2 = 1$ and the highest $u_1/u_2 = -1$. The condition $u_1/u_2 = 1$ represents the situation in which the two orbitals have equal weight in the formation of the molecular bond, the wavefunction amplitude being significant in the region between the nuclei. This is the case of homopolar covalent bonding, where the molecule formation is favored.
- If $V_2 > 0$ instead, the opposite occurs with the lowest state having $u_1/u_2 = -1$ and the highest $u_1/u_2 = 1$. If $u_1/u_2 = -1$, the wavefunction in the region between the nuclei tends to have a vanishing amplitude.

2. $H_{11} = \varepsilon_1 \neq H_{22} = \varepsilon_2$ and V_2 real. The linear system of equations reads

$$\begin{aligned} (\varepsilon_1 - \varepsilon)u_1 + V_2u_2 &= 0 \\ V_2u_1 + (\varepsilon_2 - \varepsilon)u_2 &= 0, \end{aligned} \quad (1.74)$$

yielding the eigenvalues ε and eigenvector ratios u_1/u_2

$$\begin{aligned} \varepsilon &= \frac{\varepsilon_1 + \varepsilon_2 \pm \sqrt{(\varepsilon_1 - \varepsilon_2)^2 + 4V_2^2}}{2} \\ \frac{u_1}{u_2} &= -\frac{V_2}{(\varepsilon_1 - \varepsilon)}. \end{aligned} \quad (1.75)$$

The quantities $\bar{\varepsilon} = (\varepsilon_1 + \varepsilon_2)/2$ and $V_3 = (\varepsilon_1 - \varepsilon_2)/2$ can be introduced, which are mathematically and physically sound. They indeed represent the average

value of the atomic energies and the relative electronegativity of the two atoms, respectively. In terms of the new quantities, the eigenvalues are

$$\varepsilon = \bar{\varepsilon} \pm \sqrt{V_3^2 + V_2^2}. \quad (1.76)$$

Once again, the lowest energy corresponds to the negative sign, the shift from $\bar{\varepsilon}$ being independent of the V_2 and V_3 signs. Let now look at the probability of finding the electron close to either one of the two atoms, that is the weight u_i , $i = 1, 2$ of each orbital referred to the total $u_1^2 + u_2^2$. Use of (1.74) and (1.76) yields:

$$\begin{aligned} \frac{u_1^2}{u_1^2 + u_2^2} &= \frac{1}{2} \left(1 - \frac{V_3}{\sqrt{V_3^2 + V_2^2}} \right) \\ \frac{u_2^2}{u_1^2 + u_2^2} &= \frac{1}{2} \left(1 + \frac{V_3}{\sqrt{V_3^2 + V_2^2}} \right). \end{aligned} \quad (1.77)$$

Equation (1.77) has an immediate physical meaning: the probabilities to find the shared electron close to either one of the two atoms are different and they do so by the amount $V_3/\sqrt{V_3^2 + V_2^2}$. The quantity

$$\alpha_p = \frac{V_3}{\sqrt{V_3^2 + V_2^2}} \quad (1.78)$$

represents the ionicity or bond polarity, depending on both V_2 and V_3 , including the sign of the latter. The following limiting cases are physically sound:

- The case 1 above is recovered while $|V_2| \rightarrow \infty$, and leads indeed to a purely covalent bond with $\alpha_p \rightarrow 0$
- $|V_3| \rightarrow \infty$, so that $\alpha_p \rightarrow \pm 1$ and a purely ionic bond is obtained

Once the molecular orbital for the single electron is found, the two-electron orbital has to be built up. Pauli principle requires that this be antisymmetric with respect to exchange of the two electrons. This goal is accomplished by multiplying the spatial part $\psi(\mathbf{r}_1)\psi(\mathbf{r}_2)$ of the wavefunction by the antisymmetric spin function $\chi(1, 2)$ as follows:

$$\Phi(\mathbf{r}_1, \mathbf{r}_2) = \psi(\mathbf{r}_1)\psi(\mathbf{r}_2)\chi(1, 2). \quad (1.79)$$

On the other hand, the spatial part of the orbital can be expressed as

$$\begin{aligned} \psi(\mathbf{r}_1)\psi(\mathbf{r}_2) &= u_1^2\psi_1(\mathbf{r}_1)\psi_1(\mathbf{r}_2) + u_2^2\psi_2(\mathbf{r}_1)\psi_2(\mathbf{r}_2) \\ &\quad + u_1u_2[\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2) + \psi_2(\mathbf{r}_1)\psi_1(\mathbf{r}_2)], \end{aligned} \quad (1.80)$$

with norm $u_1^2 + u_2^2$. The first line on the right-hand side represents the polar part of the orbital, with each of the two electrons being predominantly localized on the

corresponding atom with amplitude probability $u_1^2/(u_1^2 + u_2^2)$ and $u_2^2/(u_1^2 + u_2^2)$, respectively. The second line represents the purely covalent part of the orbital, the two electrons being fully shared between the two atoms, with amplitude probabilities $u_1 u_2 / (u_1^2 + u_2^2)$.

The simple case of a biatomic molecule with two sharable electrons can be generalized to more than one electron: equal atoms share a purely covalent bond, while different atoms share an orbital containing both polar and covalent parts.

Realizations of purely covalent crystals are for example diamonds, silicon, and germanium: the four valence electrons of each atom, two s and two p, get to be redistributed over four sp^3 hybridized orbitals $|h_i\rangle$:

$$\begin{aligned} |h_1\rangle &= \frac{1}{2} (|s\rangle + |p_x\rangle + |p_y\rangle + |p_z\rangle) \\ |h_2\rangle &= \frac{1}{2} (|s\rangle + |p_x\rangle - |p_y\rangle - |p_z\rangle) \\ |h_3\rangle &= \frac{1}{2} (|s\rangle - |p_x\rangle + |p_y\rangle - |p_z\rangle) \\ |h_4\rangle &= \frac{1}{2} (|s\rangle - |p_x\rangle - |p_y\rangle + |p_z\rangle) . \end{aligned} \quad (1.81)$$

The four orbitals are to all respects equivalent and highly directional, thus originating a strong interaction energy between pairs of neighbor atoms, and eventually to the tetrahedral structure already discussed in Sec. 1.2. [Table 1.11](#) lists a number of significant parameters characterizing purely covalent crystals, all of them sharing the same crystal structure.

Table 1.11 Cohesion in covalent crystals (from [3]). The values of first-neighbors distance R_0 , melting temperature T_m , and cohesive energy E_c are listed for the selected elements in the first column

| Material | R_0 (Å) | T_m | E_c (eV/atom) |
|------------|-----------|--------|-----------------|
| Diamond | 1.54 | > 3500 | 3.68 |
| Silicon | 2.35 | 1410 | 2.32 |
| Germanium | 2.48 | 937 | 1.94 |
| Tin (gray) | 2.80 | 232 | 1.56 |

[Table 1.12](#) lists the same parameters as in [Table 1.11](#) for different isoelectronic materials, that is materials which effectively share the same row of the periodic table of elements. Here, the corresponding values of ionicity α_p in (1.78) are listed. The melting temperature and the cohesive energy in covalent crystals are decreasing functions of R_0 . As well, the ionicity of isoelectronic compounds, having similar values of R_0 , increases with increasing the melting temperature and with decreasing cohesive energy.

Table 1.12 Cohesion in covalent crystals (from [3]). The values of ionicity α_p , melting temperature T_m , cohesive energy E_c are listed for the isoelectronic selected compounds in the second column, belonging to the combination of groups indicated in the first

| Group | Compounds | α_p | T_m (K) | E_c (eV/atom) |
|-------|-----------|------------|-----------|-----------------|
| IV | Ge | 0% | 937 | 1.94 |
| III-V | GaAs | 4% | 1238 | 1.63 |
| II-VI | ZnSe | 15% | 1517 | 1.24 |

Finally, the case of materials where different bonds are predominant along different directions remains to be considered. One such typical example is graphite. This is characterized by Van der Waals bonds in direction perpendicular to the hexagonal planar structure, namely between Carbon atoms lying on adjacent planes. Carbon atoms lying on the same plane are instead characterized by covalent bonds, with sp^2 hybridization of the three atomic orbitals:

$$\begin{aligned} |h_1\rangle &= \frac{1}{\sqrt{3}} (|s\rangle + |p_x\rangle + |p_y\rangle) \\ |h_2\rangle &= \frac{1}{\sqrt{3}} (|s\rangle + |p_x\rangle - |p_y\rangle) \\ |h_3\rangle &= \frac{1}{\sqrt{3}} (|s\rangle - |p_x\rangle + |p_y\rangle) . \end{aligned}$$

(1.82)

In $x - y$ plane, the three of them form three bonds along the three directions separated by a $2\pi/3$ angle, while the orbital p_z cooperates to form the bond between atoms in adjacent planes. As already seen, atoms within the same plane are separated by about 1.42 Å. The shorter distance between the atoms in adjacent planes is instead twice as much, about 3.35 Å, determining the highly exfoliation capability of graphite along the planes.

Quite a lot of materials crystallize into structures characterized by covalent and polar bonds, including inorganic, organic, and biological structures.

Quick Questions

- Q15.

Discuss whether and why Ge can be considered to be characterized by purely covalent bonding.
Answer. Ge is characterized by purely covalent bonding: indeed $V_3 = 0$ and, therefore, from (1.78) it results $\alpha_p = 0$.

1.7.5 Metallic bonds

On phenomenological grounds, metallic bonds in crystals have a well definite characterization which makes them different from all the other bonds considered so far. There, atoms or molecules do not modify in a significant way their electronic configuration once in the crystal: almost not at all in Van der Waals crystals; to some extent in ionic systems where one or more electrons have amplitude probabilities maximally transferred towards the ion with larger electronic affinity to a more relevant extent in covalent crystals, where electron wavefunctions are delocalized between two atoms, feeling their attraction. A different picture shows up for metallic compounds.

Concept

In metallic bonds the number of first neighbors n_c in the crystal structure is typically larger than the number of valence electrons of the isolated atoms. Each electron has n_c equivalent sites which it can share its wavefunction with. The latter becomes even more delocalized than in covalent bonds, further reducing its kinetic and potential energies for similar reasons.

Indeed, metallic elements are typically found while scrolling down the first two columns in the periodic table and in those next to the transition elements. As a general rule, at any fixed column of isolated elements, increasing distances between first neighbors in the crystal structure corresponds to lower cohesive energies and melting temperatures.

A simple model can be set up, which explains the main characteristics of a metal bonding. The basic concept is that valence electrons be considered as free particles, weakly interacting with each other and with the ions sitting at the lattice sites. The electrostatic interactions between electrons and nuclei and of electrons among themselves are here schematically discussed, a detailed treatment being postponed later on in this textbook.

Here is the model sketched. As discussed in Appendix 1.10 and detailed in Chap. 2, free electron wavefunctions in a box with periodic boundary conditions in a volume V behave as stationary waves characterized by quantized wavevectors \mathbf{k} : the \mathbf{k} -space volume associated to each energy state is $\Delta k = (2\pi)^3/V$ and a thin slice $d\mathbf{k}$ contains a number of states amounting to $d\mathbf{k}/\Delta k = V(2\pi)^{-3}d\mathbf{k}$. At zero temperature the occupation of the electron energy states is realized by progressively filling in all the states with increasing energy until all the N electrons have been arranged at most one each state because of Pauli principle. This procedure defines the maximum possible energy at zero temperature ε_f , named Fermi energy: this can be expressed as $\varepsilon_f = \hbar^2 k_f^2/(2m)$ in terms of the largest wavevector k_f , representing the radius of the sphere containing all the occupied states. Thus:

$$N = \frac{2V}{(2\pi)^3} \int d\mathbf{k} = \frac{2V}{(2\pi)^3} 4\pi \int_0^{k_f} k^2 dk = \frac{V}{\pi^2} \frac{k_f^3}{3}, \quad (1.83)$$

the factor of 2 on the first line accounting for the two possible spin states. Equation (1.83) leads to the fundamental relation

$$k_f = (3\pi^2 n)^{\frac{1}{3}}, \quad (1.84)$$

connecting the density of electrons n to the Fermi wavevector k_f .

Two different models can be constructed to discuss the metallic bond in crystals, according to two different approximations for the interactions with the crystal ions.

1. Consider the solid as if the ions were represented by an uniform distribution of positive charge. It can be demonstrated that simple plane waves are the wave functions which solve the Schrödinger equation for the electrons considered as free indistinguishable particles. Fundamentals and characteristics of this model are going to be discussed in Chap. 2. In essence, the total energy E_t per particle is composed of the average kinetic energy $\bar{\epsilon} = 3\epsilon_f/5$ as calculated in Appendix 1.10, and the exchange energy originated by the Pauli principle. The interaction energies are considered, as stated, negligible. The cohesive energy is thus only due to electrons, and it is

$$E_c = \frac{3}{5}\epsilon_f - \frac{3}{4\pi}e^2k_f = \frac{3}{10}\frac{\hbar^2}{m}k_f^2 - \frac{3}{4\pi}e^2k_f. \quad (1.85)$$

At equilibrium $\partial E_t / \partial k_f = 0$, and this occurs for

$$k_f = 5/(4\pi a_b),$$

with $a_b = \hbar^2/(me^2)$ being the Bohr radius. For each electron, the isotropic spherical volume $4\pi\bar{r}^3/3 = 1/n$ is on average available. This yields $\bar{r}/a_b = (4\pi/5)(9\pi/4)^{1/3} \simeq 4.82$, so that $\bar{r} \simeq 2.55 \text{ \AA}$: the sphere available to each electron has a constant average radius, independent of the geometrical structure of the material. Finally, at equilibrium one has

$$E_c = -15(32\pi^2)^{-1}(e^2/a_b) \simeq -1.3 \text{ eV/atom},$$

a value very close to the one found experimentally.

2. The electrostatic energy of interaction between the valence electrons and the ions with charge Ze can be introduced, correcting for the existence of the core electrons. In this model, the valence electrons can be thought to be uniformly distributed on a sphere of radius r_0 . The electrostatic interaction term can now be expressed as:

$$U_1 = -\alpha \frac{Z^2 e^2}{2r_0}, \quad (1.86)$$

where the factor $\alpha = 9/5$ effectively accounts for the presence of the core electrons on the given ion. The interaction among the spheres can thus be calculated, leading to an expression of the Madelung type with α_m depending on the geometric structure. In particular, it turns out that $\alpha = 1.79$ for fcc, bcc, and

hexagonal structures, whereas $\alpha = 1.67$ for diamond, and $\alpha = 1.76$ for a simple cubic lattice. Comparison with model 1 above requires an expression for r_0 in terms of k_f . This can be inferred by assuming that each atom with Z valence electrons have a sphere of radius r_0 available, so that

$$n = \frac{Z}{V} = \frac{3Z}{4\pi r_0^3},$$

and therefore

$$r_0 = \left(\frac{9\pi Z}{4} \right)^{1/3} \frac{1}{k_f}. \quad (1.87)$$

Collecting all together, the resulting total energy per site is:

$$E_c = \frac{3}{10} \frac{\hbar^2}{m} k_f^2 - \frac{3}{4\pi} e^2 k_f - \alpha \frac{Z^{5/3} e^2 k_f}{(18\pi)^{1/3}}. \quad (1.88)$$

At equilibrium $\partial E_c / \partial k_f = 0$ and one obtains:

$$k_f = \frac{1}{a_b} \left[\frac{5}{4\pi} + \frac{5}{6} \alpha Z^2 \left(\frac{4}{9\pi Z} \right)^{\frac{1}{3}} \right]. \quad (1.89)$$

Note that with respect to case 1 above, k_f here is larger and $r_0 = (9\pi Z/4)^{1/3} / k_f$ correspondingly smaller. This finding is consistent with the present model setting, where a collective term is introduced, favoring the atom packing. As well, the values of r_0 and of the cohesive energy calculated within this model are somehow compatible with those experimentally observed. However, strictly speaking they are not close enough, so that the discussed model is not yet fully realistic.

By neglecting the last term in the right-hand side of (1.89), the compressibility modulus of a metal is found to be $B_0 = (2/3)n\epsilon_f$, indicating that the ground-state electronic configuration is directly linked to the mechanical properties. In particular, it turns out that an ideal metal with high density and large Fermi energy, can be hardly compressed. This is easily explained from (1.85): compression of the sample would rise the total kinetic energy as $1/r_0^2$, faster than the energy gain associated to the enhanced Coulomb attraction, that scales instead as $1/r_0$.

The above model clarifies why metallically bonded solids tend to crystallize into closed packed structures such as hcp, fcc, bcc, according to what explained in Sec. 1.3.1: such structures maximize space filling and coordination number, due to the isotropic nature of the Coulomb attraction between the electron cloud and the positively charged ions.

1.7.6 Crystal bonding and periodic table

Considered the discussion conducted so far on the bonding models, a question arises: can one guess the bonding model for a given material, once the atomic species involved in the bonds formation are known? While a definite answer to this question cannot be elaborated, a few rules of thumb can be sketched to guide the answer. These rules stem from the knowledge of some chemical properties of the individual atoms, that are now being given.

Definition

Metallicity can be defined as the aptitude of an atom to donate electrons in metallic or ionic bonds.

Metallic aptitudes of elements increase while moving from right to left or from top to bottom in the periodic table, that is while the capability of electron binding decreases.

Definition

Electronegativity is the opposite of metallicity: it is the aptitude of an atom to attract an electron.

Electronegativity follows the opposite trend of metallicity and thus increases while moving from left to right and from bottom to top of the periodic table.

The above definitions favor the distinction between metallic and non-metallic elements. Next, the following rules can be set: (i) the bond between a metallic and a non-metallic element can be either ionic or covalent, depending on their electronegativity difference; (ii) metallic elements form metal-like bonded solids; (iii) non-metallic elements form covalently bonded solids. As a matter of fact however, in some cases the metallic or non-metallic character of a given chemical element cannot be established with sufficient evidence. Therefore the above rules must be taken just as general guidelines.

As a final remark of this section, most of the material properties are strongly dependent on the bond nature. Metals do usually show high thermal and electronic conductivity, high reflectivity, low to intermediate melting temperatures, large ductility, that is ability to deform under tensile or compressive stress. On the other hand, ionic-like bonded solids are thermal and electronic insulators, usually appear as transparent and colorless, have intermediate to high melting temperatures, are brittle and are soluble in polar solvents. Covalently bonded solids are typically insulators, non reflective, brittle, with high melting points. Finally, solids where van der Waals interactions do play significant role tend to be insulators, transparent, soft, soluble and with low melting temperatures.

The listed mechanical properties are the subject of the next section. The listed transport and optical properties such as reflectivity, electrical and thermal conductivity, are treated in Chaps. 4-5.

1.8 Mechanical properties of materials and crystal structures

Concept

Crystal structure and bonding nature significantly determine the mechanical properties of materials.

In technological applications, materials are required to have best performances with respect to either mechanical or electronic properties, or both.

Concept

Single elements or compounds hardly have such high performances, so that it is often needed to engineer on purpose structures by means of suited physical and chemical processes.

1.8.1 Crystalline materials

This section is devoted to a schematic discussion of mechanical performances by means of selected examples. The main concepts useful to describe a mechanical deformation of a body are firstly introduced.

Definition

The mechanical stress is a measure of the force required to cause a particular deformation.

In practice, it measures the average force per unit area of a surface within the body. Normal and shear-type stresses are usually classified. Normal stress, often labeled with the symbol σ , arises whenever either tensile or compressive forces act against each other. A typical example is an axially loaded body, such as a bar subjected to tension or compression due to a force passing through its center. This is depicted in Figs. 1.39a-b. Tensile or compressive forces always act along the same axis in opposite directions. If two forces act as a torque, that is in opposite directions and at some distance between them, a shear stress shows up instead. This is displayed in Fig. 1.39c, the shear stress being often indicated with the symbol τ . As it is clear from Fig. 1.39, whenever a force acts on a cross sectional area of a body, the normal stress arises from the force component F_n normal to the cross section A , that is defined as $\sigma = F_n/A$. Shear stress arises from the parallel component F_p instead, and is defined as $\tau = F_p/A$.

Definition

The strain is defined as the normalized measure of the deformation of the material induced by the applied force.

As schematically depicted in Fig. 1.40, in simple cases the strain is calculated as the ratio of the deformation to the original length, that is $\varepsilon = \Delta L/L$.

Sometimes hard materials are wanted.

Definition

Hardness qualifies the resistance that materials offer to penetration and ablation. Therefore, it is strongly correlated to specific type of bonds.

As expected, covalent crystals are especially hard materials, covalent bonds being highly directional. A typically hard material is diamond, indeed. The larger the energy associated to the covalent bond, the higher the degree of hardness. For this reason elements from the same column of the periodic table as diamond, such as silicon and germanium, result to be less hard: their corresponding larger ion sizes determine smaller bond energies.

Sometimes plastic materials are needed.

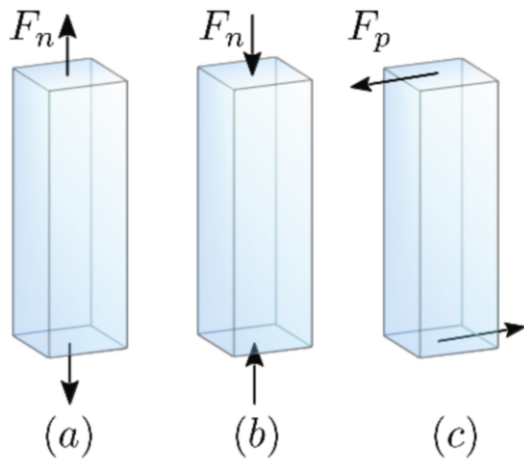


Fig. 1.39 Deformations in solids. Schematic representation of (a) tensile, (b) compressive, and (c) shear forces

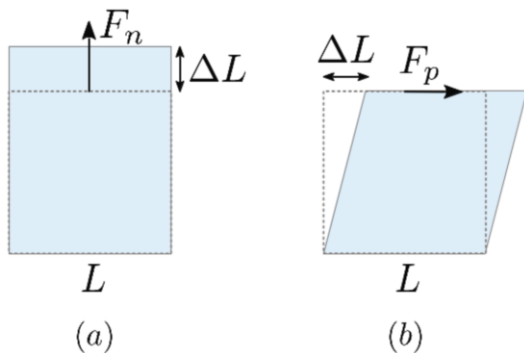


Fig. 1.40 Deformation in solids. Schematic representation of a deformation in the case of (a) normal and (b) shear stress

Definition

Plasticity qualifies the ease that materials show to be stressed and to permanently change shape as a result of a strain beyond a given limit, called yield.

A stress-strain picture is of help here and shown in Fig. 1.41. Well before the yield limit, the same material can generally show

Definition

Elasticity properties: elasticity is the capability of progressively changing shape under increasing but tiny stress, and releasing back the strain once the stress is removed. Thus, returning back to initial shape and size.

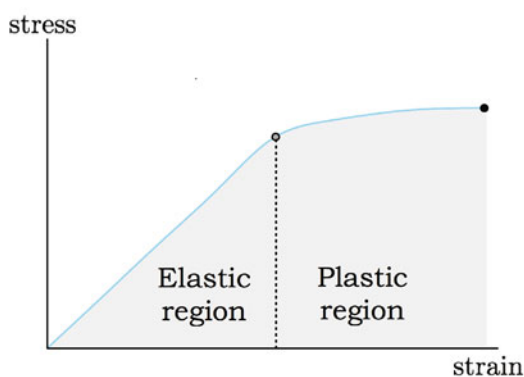
Definition

Resilience is instead the ability that materials have to store energy during an elastic strain, in practice the area under the elastic region in the stress-strain graph.

Beyond plastic capabilities, the material is fragile, and susceptible of being broken into pieces.

The elasticity region is schematically depicted in Fig. 1.41. If the applied stress gets too large values, the proportionality relationship between strain and stress ceases to be valid. In this plastic region, the material is permanently deformed by the stress. The maximum value of the stress is the fracture limit, beyond which fractures occur indeed in the material.

Fig. 1.41 Deformation in solids. An oversimplified schematic representation of the stress versus strain curve. The filled grey and black dots represent the elasticity limit (yield) and the fracture point, respectively



Plasticity can be originated by bonding nature as well as structural characteristics. Many molecular carbon-based compounds, such as polymers, are highly plastic

due to their coiled and entangled structures. Selected metals, such as gold and copper, are especially plastic mainly because of bonding nature.

In fact, metals are ductile and malleable, meaning that they can be deformed permanently under different type of strain, before fragility shows up.

Definition

Ductility is the ease offered by the material to be deformed after stretching, without remaining damaged in its structure or mechanical resistance. Malleability refers to the same kind of response, but to a compression strain instead than a stretching.

Though similar in the kind of mechanical response, malleability and ductility do not get precisely along: in fact malleability is useful for rolling materials into sheets and ductility to stretch them into wires.

Concept

Whatever the stress be, the response is in both cases related to how much easy it is to shift atoms from their lattice positions.

In metallic crystals ions are embedded in an electron sea, and slipping of atoms or atom layers past each other is somewhat easy. The same occurs with materials showing defects like dislocations, again quite common in metals, where an atomic plane can be imperfectly ended. For opposite reasons, covalent crystals with their highly directional bonds are expected to be neither ductile nor malleable, more easily tending to shatter under a stress. Ionic crystals are generally not ductile and even more inclined to shatter, as they need to break up so many bonds to shift atoms around. And should a shift occur, this would line up ions of the same sign, raising up electrostatic repulsions.

Concept

In general, any type of defect able to contrasting the effectiveness of dislocations would reduce ductility and effectively increase hardness.

This can be the case of interstitial impurities, namely atoms of different type penetrated into and added to the hosting crystal structure. It can also be the case of grain boundaries, that is planar defects separating regions with different crystalline orientation in a polycrystalline solid.

Finally, a common requirement is that the material be chemically inert, especially with respect to air agents. Iron for example is easily oxidized in contact with air, becoming rusty.

Concept

One way to improve the material ability to be chemically inert without significantly affecting their mechanical properties is to introduce substitutional impurities: an atom of different type substitutes indeed the original atom of the hosting crystal when a vacancy defect should occur. Substitutional impurities generally have similar sizes than the hosting atoms, and thus are amenable to modify the chemical properties of the material while leaving the mechanical ones essentially unchanged.

One example is steel, resulting from Fe subjected to work-hardening, thermal cycles and addition of small interstitial impurities. Work-hardening is a process aimed to increase the number of dislocations up to make them so correlated that they cannot move. Thermal cycles are good to reduce grain sizes. Typical interstitial impurities are C, H, and N atoms. In addition, Ni and Cr atoms are substituted to Fe atoms to make it less exposed to rusting processes. A whole branch of science and technology is exclusively devoted to chemical preservation to maintain correct device working conditions.

Quick Questions

Q16. You are a famous cook and as long a famous physicist. You need to choose a quite good knife. Are you going to construct it from iron or steel, or else... [1]?

Answer. Iron to have a softer but also less fragile knife, steel to have it harder but unfortunately more fragile and, if containing substitutional impurities, rust-free. Perhaps something hardest and not so fragile? Sometimes different parts of knives are made of different materials.

To conclude, [Table 1.13](#) summarizes selected properties among those discussed so far, relating them to the bond type.

Table 1.13 Selected properties of solids are listed as a function of the bond type

| Van der Waals | Ionic | Covalent | Metallic |
|-------------------|--------------------|-------------------------|------------------------|
| Ne, Ar, Kr, Xe | NaCl, CsCl, ZnS | C (diamond), Ge, Si | Fe, Cu, Ag, Na |
| soft and brittle | hard and brittle | very hard | variable hardness |
| non conducting | non conducting | usually not conducting | conducting |
| low melting point | high melting point | very high melting point | variable melting point |

1.8.2 *Non-crystalline materials*

At the beginning of this Chapter the materials have been classified as crystals, polycrystals and liquids according to the spatial behavior of the correlation function. Glass state has been introduced as an example of amorphous atom arrangement, plastics and polymers as materials that are partly in crystalline and partly in amorphous state. For the sake of completeness, the behavior of such systems under mechanical stress is here very briefly and qualitatively discussed [1].

Glasses. Examples of glasses are silica, made after SiO_2 , other glass formers are GeO_2 , P_2O_5 , and B_2O_3 compounds. As a general consideration, glasses respond to shear stress more like a solid than a liquid would do: in glass state atoms cannot in fact slide one across the other. The response to temperature gradients is as well solid-like: glass shrinks when cooled, but less than a liquid would do.

Glass as a material is one everyday-life example of glass state: it is made after crystalline silica, SiO_2 , that is characterized by a very large viscosity at its freezing temperature, after very rapid cooling. Quartz is an example of crystalline silica, with trigonal unit cell. In the cell are 7 SiO_4 units with silicon at the center of a tetrahedron and oxygen atoms arranging at the apex, each oxygen bridging between two Si by means of highly directional covalent bonds. When melting, such a very strong crystal arrangement transforms into a network of interconnecting rings made of alternating Si and O, each Si sharing with three O of three different rings and a fourth O out of the ring plane. The number of atoms in each ring is in fact variable. Nevertheless, covalent bonding and ring arrangement make the melting temperature of glassy silica as high as 1600°C or higher, that is above iron and steel melting temperatures. For this reason silica are often modified with addition of soda, Na_2O , that lowers the melting temperature down to 793°C : in essence, Na gives its electron away to oxygen atoms in silica, which thus need only one more electron and bind to only one Si instead than bridging between two. The presence of Na ions makes silica with soda too much soluble in water: while this is useful to paint surfaces and in laundry detergents, it is no way good as a glass since it easily breaks and scratches. To recover a better solid glass, lime, or CaO , is often added because it is an ionic solid that is not soluble in water. Windows and bottles are usually made of silica modified with soda and lime.

Strengthening of glasses is performed by tempering processes. In essence, surfaces are cooled more rapidly than the bulk, down to the transition temperature for glass formation: the bulk, that remains liquid, shrinks after subsequent cooling more than the glassy surface, that is therefore subjected to huge compressive stresses. Cracking of tempered glasses requires large forces than in the case of normal glasses: however, when the crack occurs, the tempered glass crumbles into tiny pieces, it does not just break. For this reason, car front glasses for example, or bullet-proof glasses are made by alternating layers of glass and plastic sheets, the latter working to safely keep the glass together in a crack event.

Addition of other modifiers can engineer the optical properties of glasses. Though optical properties are the subject of Chap. 5, at the present qualitative level it can be

evidenced that addition of lead oxide is often used to change the glass index of refraction and also to make it efficient absorbers of X -radiation, while colored glasses are obtained by adding metal ions, the color depending on the energy difference between the empty electronic states, where electrons can jump after absorbing light, and the ground state.

Plastics and polymers. Natural polymers are for example natural rubber, cellulose, hair, and gelatin, DNA. Popular synthetic polymers are polyethylene, celluloid, and Plexiglas®, Nylon®, Teflon®, Kevlar®. As a general consideration, the long and entangled chain structure of polymers, along with the presence of hydrogen bonding, gives them robust mechanical characteristics [1]. In fact, materials exist that are in form of chains arranged in an entirely crystalline structure: natural (animal or vegetable) or synthetic wax is one example, a type of lipid with a great variety of mainly alkanes (such as the simplest methane CH_4) but also various kinds of esters (such as ethyl acetate). Comparing with wax, polymer-based plastics do not crumble when bent, because of the presence of the amorphous part. Plastics are stronger under tensile stresses: polymer chains are much longer than in wax. They are also entangled, while in wax they are arranged in crystalline structures held together by van der Waals forces. Polymer chains can untangle only by reptation, that is by sliding one out of the other: when subjected to increasing temperature for example, polymers pass through rubber-like, then rubber-like flow regimes, and eventually to liquid-like flow regime at higher temperature.

Natural rubber is in essence made of chains of isoprene units, that is with brutal chemical formula C_5H_8 . Hair is made of chains of protein molecules that are in turn made on average of something like 20 amino acids: glycine, cysteine, keratin are examples. Gelatin is made of billions of long thin protein molecules, partly in rigid crystalline form and partly in sticky amorphous form, so that when mixed to water, water molecules can be trapped within the fibers and make the substance solid but soft and elastic. Polyethylene basically is a chain of $(\text{C}_2\text{H}_4)_n$ units. Cellulose is composed of glucose units connected by a C-O-C unit, a glucose unit being a ring with five C and one O atoms, the C atoms bonded to four other atoms including C, H and O: overall thus, cellulose is very stiff, the only way to break it into smaller sugar molecules (and digest it!) being by action of bacteria and protozoans such as those that ruminants have. At variance with cellulose, Nitrocellulose or celluloid can be plasticized, though it becomes in turns flammable and explosive. Plexiglas® is made by long chains of methyl methacrylate units with brutal chemical formula $\text{C}_5\text{H}_8\text{O}_2$ units: this strong bonding structure makes Plexiglas® hard and durable, solid below 105°C . Nylon® is one more example of hydrocarbon-based chain, characterized by alternating base and acid units hydrogen-bonded to each other, each containing a number of hydrocarbon pieces and terminated by N atoms: the shorter each unit, the larger is the number of hydrogen bonds and thus the Nylon® stiffness. Kevlar® follows the same concept of Nylon®, but with aromatic rings instead of hydrocarbon chains, giving the material its extraordinarily large tensile strength. Finally Teflon® is conceptually similar to polyethylene, but with the carbon chains

surrounded by F: Teflon® is in fact very reluctant to undergo chemical reactions, and is indeed used for example to coat pans.

Summary: concepts, tools, and procedures to know

Concepts and Tools

- **Different types of ordering exist, such as crystalline, polycrystalline and amorphous. They can be represented via the concepts of distribution function and of radial pair-correlation function. The latter is the probability of finding a particle at distance r from another selected one.**
- **Crystal structures can be described by Bravais lattices: lattices of points arranged in regular and periodic manner, which look always the same from whatever perspective and are left unchanged after a translation by a lattice vector \mathbf{R} .**
- A real crystal is identified by a primitive cell and a base.
- Any symmetry operation can be obtained as combination of a point-group and a translation operation.
- Lattice-symmetry operations limit the kind of allowed pavings: for example, no pentagonal paving is allowed.
- FCC and close-packed hexagonal structures are most common structures, the crystal structure depending also on temperature and pressure.
- **Reciprocal lattice and related tools are useful to assist the comprehension of experiments to determine crystal structure.**
- At any given direction a one-to-one correspondence exists between vectors of the reciprocal lattice with smallest size and families of lattice planes, that are then usefully labeled via Miller indexes.
- **The investigation of a system geometrical structure requires particle or radiation-like probes of suited nature, wavelength and energies. This is usually conducted by means of elastic scattering of electromagnetic waves at suited frequencies: charged particles subjected to these waves, in turn behave as antennas, that is emitting sources of radiation at the same frequencies. The investigation of excitations instead better requires inelastic scattering.**
- In elastic scattering, the directions of maximum scattered radiation or particles identify the reciprocal lattice vectors and therefore the structure of the direct Bravais lattice. The intensity modulation of scattered waves carries information on the Bravais-lattice arrangement, configuration of the base, and nature of bonds, each information being represented via a different (geometrical or atomic) structure factor.
- The scattered radiation intensity is highest whenever $\Delta \mathbf{k}$ fits a vector \mathbf{G} of the reciprocal lattice.

- Bragg's law (1.44) is a restatement of crystal periodicity, with no trace of details on atomic positions and basis composition. Bragg reflection requires wavelengths $\lambda \leq 2d$, thus much smaller than those of visible light.
- Scaling of scattered intensity with the number N or N^2 of sources indicates random or regular arrangements, respectively, and is a useful concept in studying matrix-embedded nanostructures.
- STM is based on the measure of a tunnel current. AFM on the measure of cantilever deflection by atomic tip-surface forces.
- **Cohesive energy and compressibility are indicators of stability in crystal formation.**
- Interaction potentials in crystals generally include a repulsive part responsible for matter impenetrability, and an attractive part responsible for bonding and thus materials cohesion.
- Several mechanisms concur in determining the cohesion properties of materials: Van der Waals-like bonds based on induced dipolar Lennard-Jones interactions; ionic bonds based on electrostatic forces; covalent bonds where electrons are shared in orbitals with amplitude probability delocalized all over the molecule; and metallic bonds with electron wavefunction delocalized over all sites. These can be calculated from first principles or from simpler phenomenological few parameters models, to be compared with measurements of cohesive energy, compressibility, ionicity, and melting temperatures.
- **Crystal structure and bonding determine the mechanical properties of materials. Understanding of the underlying microscopic mechanisms allows on-purpose material engineering.**
- **Stress-strain graphs define elastic, plastic, and breaking up regions. Hard materials are often also fragile.**
- Microscopic mechanisms underlying plasticity are generally connected with the ease to move atoms from their lattice positions. In general then, ductility can be reduced and hardness increased by defects contrasting dislocations effectiveness, such as interstitial impurities or grain boundaries.
- Substitutional impurities work to modify the chemical properties of materials leaving the structural ones essentially unchanged.

Procedures

- Construction of a Bravais lattice.
 - Construction of the Wigner-Seitz cell.
 - Determination of a family of lattice planes from Miller indexes.
 - Connection between theory and experiments in determining the crystal structure by means of radiation or particle scattering.
 - Calculation of the Madelung constant.
-

1.9 Appendix. Calculation of the transmission coefficient through a one-dimensional potential barrier

Let an incoming particle of mass m and energy E be moving along the z -axis towards a barrier of height U_0 and thickness d . The Schrödinger equation for the wavefunction of this particle is:

$$-\frac{\hbar^2}{2m}\psi''(z) + U(z)\psi(z) = E\psi(z), \quad (1.90)$$

where the $''$ indicates the second derivative with respect to z , and

$$U(z) = \begin{cases} 0 & \text{for } z < 0 \\ U_0 & \text{for } 0 < z < d \\ 0 & \text{for } z > d \end{cases}. \quad (1.91)$$

Equation (1.90) has to be solved in the regions $L \equiv z < 0$, $C \equiv 0 < z < d$, and $R \equiv z > d$. The particle is supposed to come from $z \rightarrow -\infty$, and be in part reflected by and in part transmitted through the potential barrier. The particle wavefunctions Ψ_L on the left, Ψ_C on the center, and Ψ_R on the right regions can be written in terms of the amplitudes r of the reflected and t of the transmitted waves, in the form:

$$\begin{aligned} \Psi_L(z) &= e^{ik_0z} + re^{-ik_0z} \\ \Psi_C(z) &= B_re^{k_1z} + B_le^{-k_1z} \\ \Psi_R(z) &= te^{ik_0z} \end{aligned} \quad (1.92)$$

where $k_0 = \sqrt{2mE/\hbar^2}$ and $k_1 = \sqrt{2m(U_0 - E)/\hbar^2}$, while r , B_r , B_l , and t are constants to be determined. The reflection and transmission coefficients are expressed as $R = |r|^2$ and $T = |t|^2$. Equation (1.92) contains four unknowns, that are r , t , B_r , and B_l . Continuity conditions of the wavefunction and of its derivative at barrier edges $z = 0$ and $z = d$ provide 4 conditions for the four unknowns r , t , B_r , B_l . In particular it is obtained for the coefficient t

$$t = e^{-ik_0d} \frac{(a_+^2 - a_-^2)}{a_+^2 e^{-k_1d} - a_-^2 e^{k_1d}}. \quad (1.93)$$

with

$$a_{\pm} = 1 \pm i \frac{k_0}{k_1}. \quad (1.94)$$

Equation (1.93) looks simpler in the limit $k_1d \gg 1$, when

$$T = \left| \frac{a_+^2 - a_-^2}{a_-^2} \right|^2 e^{-2k_1d} = \left| 1 - \frac{a_+^2}{a_-^2} \right|^2 e^{-2k_1d}. \quad (1.95)$$

Since a_+ and a_- share the same modulus, setting $\alpha = k_0/k_1$ and $\tan(\varphi) = 2\alpha/(1 - \alpha^2)$, one obtains

$$\left| 1 - \frac{a_+^2}{a_-^2} \right|^2 = 4 \frac{4\alpha^2}{(1 + \alpha^2)^2} . \quad (1.96)$$

Eventually, one has

$$T = \frac{16 \left(\frac{k_0}{k_1} \right)^2}{\left[1 + \left(\frac{k_0}{k_1} \right)^2 \right]^2} e^{-2k_1 d} , \quad (1.97)$$

which is the desired expression.

The limit $k_1 d \ll 1$ recovers the expected situation of total transmission. Indeed, one has $t \simeq e^{-ik_0 d}$ and thus $T = 1$.

1.10 Appendix. Energy and wavenumber of free particles

Begin with the one-dimensional case first. Consider an electron with mass m , which is moving along the x -axis on an L -long segment. The electron wavefunction is

$$\varphi(x) = \frac{1}{\sqrt{L}} e^{ik_x x} . \quad (1.98)$$

The corresponding electron energy is

$$\varepsilon(k_x) = \frac{\hbar^2 k_x^2}{2m} \quad (1.99)$$

in terms of the allowed wavenumbers k_x . These in turn are fixed by the boundary conditions. Let us assume that the latter be given by $\varphi(x) = \varphi(x + L)$. In this case, one finds

$$k_x = \frac{2\pi}{L} n_x \quad (1.100)$$

with n_x an integer number. If now the motion occurs inside a parallelepiped with sides L_1 , L_2 and L_3 , the results (1.98)–(1.100) above are modified into:

$$\begin{aligned} \varphi(x, y, z) &= \frac{1}{\sqrt{V}} e^{i\mathbf{k} \cdot \mathbf{r}} \quad \text{with } V = L_1 L_2 L_3 \text{ and } \mathbf{r} \equiv (x, y, z) \\ \varepsilon(k) &= \frac{\hbar^2}{2m} k^2 \\ \mathbf{k} &\equiv (k_x, k_y, k_z) \text{ where } k_x = \frac{2\pi}{L_1} n_x, k_y = \frac{2\pi}{L_2} n_y, k_z = \frac{2\pi}{L_3} n_z . \end{aligned} \quad (1.101)$$

Since the wavenumbers composing the vector \mathbf{k} are quantized, the volume associated to each energy state is $\Delta k = (2\pi)^3 / V$. As a consequence, the volume $d\mathbf{k}$ contains a number of states amounting to

$$dN_s = \frac{d\mathbf{k}}{\Delta k} = \frac{V}{(2\pi)^3} d\mathbf{k} . \quad (1.102)$$

At zero temperature the occupation of the energy states is realized by progressively filling all the states with increasing energy up to the condition to have arranged all the N electrons, namely:

$$N = \frac{2V}{(2\pi)^3} \int d\mathbf{k} = \frac{2V}{(2\pi)^3} 4\pi \int_0^{k_f} k^2 dk = \frac{V}{\pi^2} \frac{k_f^3}{3} . \quad (1.103)$$

The factor of 2 on the first line of (1.103) accounts for the two possible spin states. The so-called Fermi wavevector k_f on the second line instead represents the radius of the sphere containing all the occupied states. Equation (1.103) leads to the fundamental relation

$$k_f = (3\pi^2 n)^{\frac{1}{3}} , \quad (1.104)$$

connecting the density of electrons n to the Fermi wavevector. The last occupied state has energy

$$\varepsilon_f = \frac{\hbar^2}{2m} k_f^2 = \frac{\hbar^2}{2m} (3\pi^2 n)^{\frac{2}{3}} , \quad (1.105)$$

that is the maximum possible energy at zero temperature, named Fermi energy. Finally, the average energy per particle turns out to be:

$$\bar{\varepsilon} = \frac{1}{N} \frac{V}{4\pi^3} \frac{\hbar^2}{2m} \int k^2 d\mathbf{k} = \frac{1}{n} \frac{\hbar^2}{2m} \frac{1}{\pi^2} \int_0^{k_f} k^4 dk = \frac{\hbar^2}{2m} \frac{1}{n\pi^2} \frac{k_f^5}{5} = \frac{3}{5} \varepsilon_f . \quad (1.106)$$

Problems with solutions

1.1. Consider the close-packed hard-sphere model. Demonstrate that the ideal ratio c/a for the close-packed hexagonal structure is $\sqrt{8/3}$.

Solution. Consider the right-angled triangle with hypotenuse a , one cathetus $c/2$, and the other cathetus $2m/3$, where $m = \sqrt{3}a/2$ is the median of the equilateral triangle with side a . Pythagoras theorem applied to this right-angled triangle states that

$$\frac{c^2}{4} = a^2 - \frac{a^2}{3} = \frac{2}{3} a^2 ,$$

which the result is easily obtained from.

1.2. Consider a body-centered cubic lattice. Determine the families of lattice planes corresponding to Miller indexes (a) $(1, 1, 1)$ and (b) $(2, 3, 1)$.

Solution. In case (a), one has $\mathbf{G}_0 = \mathbf{b}_1 + \mathbf{b}_2 + \mathbf{b}_3 = 4\pi a^{-1}(\hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}})$ and $G_0 = 4\pi\sqrt{3}/a$, $d = 2\pi/G_0 = a/(2\sqrt{3})$. A lattice plane of the indicated family is represented by the equation $\mathbf{R} \cdot \mathbf{G}_0 = A$ with A a constant and $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$.

This leads to $\mathbf{R} \cdot \mathbf{G}_0 = 2\pi(n_1 + n_2 + n_3)$. The integer numbers n_1 – n_3 are all those that can be chosen in a way that $n_1 + n_2 + n_3 = n$ with n fixed.

In case (b), one has $\mathbf{G}_0 = 2\mathbf{b}_1 + 3\mathbf{b}_2 + \mathbf{b}_3$ and $G_0 = 20\pi/(\sqrt{2}a)$, $d = a\sqrt{2}/10$. A lattice plane of the indicated family is represented by the equation $\mathbf{R} \cdot \mathbf{G}_0 = A$ with A a constant. One finds $\mathbf{R} \cdot \mathbf{G}_0 = 2\pi(2n_1 + 3n_2 + n_3)$. The integer numbers n_1 – n_3 are all those that can be chosen in a way that $2n_1 + 3n_2 + n_3 = n$ with n fixed.

1.3. Consider the structure of diamond or silicon. Demonstrate that the first neighbors to the atom in $(0,0,0)$, are located at $(a/4, a/4, a/4)$, $(a/4, -a/4, -a/4)$, $(-a/4, a/4, -a/4)$, and $(-a/4, -a/4, a/4)$, and that the four of them are at the apex of a tetrahedron.

Solution. As already seen, diamond and silicon crystallize onto a face-centered cubit lattice. In this structure, the positions that are first neighbor to $(0,0,0)$ are: $(\pm a/2, \pm a/2, 0)$, $(\pm a/2, 0, \pm a/2)$, and $(0, \pm a/2, \pm a/2)$. The result is demonstrated by first selecting suited signs among the various \pm possibilities, and then summing to them the vector $(a/4, a/4, a/4)$. All the resulting positions are at distance $d = \frac{a}{4}\sqrt{3}$ from the origin, and form the angle φ such that $\cos(\varphi) = -1/3$. By construction, it is easily seen that the four atoms are at the vertex of a tetrahedron.

1.4. In the text the geometrical structure factor of the diamond has been discussed: there is the formation of the covalent bonds between each atom with the four first neighbors. In the conventional cubic cell the bond occurs between the atoms at $(0,0,0)$ and $(a/4, a/4, a/4)$. This fact can be effectively mimicked after locating a tiny charge at $(a/8, a/8, a/8)$, which an atomic structure factor f' corresponds to. Under these conditions, calculate the additional modulation of the scattered intensity of X-radiation, that is expected at the detector.

Solution. Expression (1.42) is now modified into

$$F_G = \left| 1 + e^{i\pi(r_1+r_2)} + e^{i\pi(r_1+r_3)} + e^{i\pi(r_2+r_3)} \right|^2 \left| 1 + e^{i\frac{\pi}{2}(r_1+r_2+r_3)} \right|^2 \times \left| f + f' e^{i\frac{\pi}{4}(r_1+r_2+r_3)} \right|^2. \quad (1.107)$$

The points satisfying the condition $r_1 + r_2 + r_3 = 4n$ determine a modulation described by the additional factor

$$\begin{aligned} &|f + f'|^2 \quad \text{if } n \text{ is even} \\ &|f - f'|^2 \quad \text{if } n \text{ is odd.} \end{aligned}$$

If instead r_1 , r_2 , and r_3 have odd-definite parity and f and f' are real, the additional factor is:

$$\left[f + f' \cos\left(\frac{\pi}{4}(r_1 + r_2 + r_3)\right) \right]^2 + \left[f' \sin\left(\frac{\pi}{4}(r_1 + r_2 + r_3)\right) \right]^2.$$

If f and f' are complex number, the additional factor can be written in a similar way.

1.5. One wants to demonstrate that the size of the reciprocal lattice vectors G depends on the r_1 , r_2 , and r_3 numbers according to a form that is characteristic of each lattice, as if they were a lattice fingerprint. Demonstrate this statement for the particular case of cubic lattices (see Eq. (1.20)).

Solution. In the simple cubic lattice, the reciprocal lattice vectors G are generally represented by

$$\mathbf{G} = r_1 \mathbf{b}_1 + r_2 \mathbf{b}_2 + r_3 \mathbf{b}_3 ,$$

with \mathbf{b}_1 , \mathbf{b}_2 and \mathbf{b}_3 given by (1.27). The modulus of G is:

$$G = \sqrt{\mathbf{G} \cdot \mathbf{G}} = \frac{2\pi}{a} \sqrt{r_1^2 + r_2^2 + r_3^2} .$$

In the case of a body-centered cubic lattice, the \mathbf{b}_i are given by (1.28) and consequently

$$G = \frac{2\pi}{a} \sqrt{2(r_1^2 + r_2^2 + r_3^2 + r_1 r_2 + r_1 r_3 + r_2 r_3)} .$$

In the case of a face-centered cubic lattice, the \mathbf{b}_i are given by (1.29) and consequently

$$G = \frac{2\pi}{a} \sqrt{3(r_1^2 + r_2^2 + r_3^2) - 2(r_1 r_2 + r_1 r_3 + r_2 r_3)} .$$

Similar calculations can be conducted for other types of lattices, each time finding a different dependence of the size of G from r_1 , r_2 , and r_3 . This result is very useful, since this dependence can be used as a lattice fingerprint in contexts where lattice structures are to be investigated and recognized.

1.6. Consider a material crystallizing into a simple cubic lattice with a two atoms base, the atoms being located at $x_1 = x_2 = x_3 = 0$ and $x_1 = x_2 = x_3 = a/2$. Calculate the intensity of elastically scattered X -radiation, that is expected at a far-away detector.

Solution. The geometrical structure factor corresponding to a conventional cell with a base of two atoms located as in the problem text, is given by:

$$F_G = f_1(\mathbf{G}) + f_2(\mathbf{G})e^{i\pi(r_1+r_2+r_3)} . \quad (1.108)$$

If the two atoms were of equal species, then $f_1 = f_2$ and (1.108) represents the geometrical structure factor of an element crystallizing into a body-centered lattice structure. If instead the two atoms were of different species, then $f_1 \neq f_2$ and (1.108) refers to compounds crystallizing in a CsCl-like structure. The F_G factor dictates the modulation of the scattered intensity along the different directions. In particular:

$$\begin{aligned} F_G &= |f_1 - f_2|^2 & \text{if } r_1 + r_2 + r_3 = 2p + 1 \\ F_G &= |f_1 + f_2|^2 & \text{if } r_1 + r_2 + r_3 = 2p . \end{aligned}$$

In addition, f_1 and f_2 depend on G : if the sum $r_1 + r_2 + r_3$ has a definite parity, then the scattered intensities are not equal, and are instead modulated according to the dependence of G on the atomic structure factors f_1 and f_2 .

1.7. Consider a total charge Q uniformly distributed with volume density ρ on a sphere of radius R_0 . Demonstrate that the electric field at the detector is proportional to

$$F(qR_0) = \frac{3Q}{qR_0} \left[\frac{\sin(qR_0)}{(qR_0)^2} - \frac{\cos(qR_0)}{qR_0} \right].$$

Solution. Consider the expression (1.36) as the starting step, with $q = |\Delta \mathbf{k}|$. Then one has:

$$\begin{aligned} E &\propto \int_V d\mathbf{r} \rho e^{-i\Delta \mathbf{k} \cdot \mathbf{r}} = \int_0^{R_0} 2\pi r^2 dr \rho \int_0^\pi e^{-iqr \cos(\theta)} \sin(\theta) d\theta \\ &= \frac{4\pi}{q} \rho \int_0^{R_0} r dr \sin(qr) = 4\pi R_0^3 \rho \frac{1}{qR_0} \left[\frac{\sin(qR_0)}{(qR_0)^2} - \frac{\cos(qR_0)}{qR_0} \right]. \end{aligned} \quad (1.109)$$

This result is identical to that in (1.46).

1.8. Derive the Hamiltonian (1.52) from the Hamiltonian (1.51).

Solution. Since

$$\sum_{i,j=1}^N \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} = \sum_{i,j=1}^N \frac{e^2}{|\mathbf{R}_0 + \xi_i - \mathbf{R}_1 - \eta_j|}, \quad (1.110)$$

$$-\sum_{j=1}^N \frac{Ze^2}{|\mathbf{r}_j - \mathbf{R}_0|} = -\sum_{j=1}^N \frac{Ze^2}{|\mathbf{R}_0 - \mathbf{R}_1 - \eta_j|}, \quad (1.111)$$

$$-\sum_{i=1}^N \frac{Ze^2}{|\mathbf{r}_i - \mathbf{R}_1|} = -\sum_{i=1}^N \frac{Ze^2}{|\mathbf{R}_0 + \xi_i - \mathbf{R}_1|}, \quad (1.112)$$

expanding to second order in ξ_i , η_j , and $\xi_i - \eta_j$ all the above terms and taking into account the repulsive term between the nuclei $(Ze)^2/|\mathbf{R}_1 - \mathbf{R}_0|$, one obtains (1.52).

1.9. Let Q be a charge uniformly distributed on the surface of a sphere of radius R . 1) Calculate the structure factor for this system. 2) Repeat the calculation for the case in which the charge be uniformly distributed in the whole sphere.

Solution. The charge distribution is uniform. The structure factor $F(G)$ is calculated by performing the integral $\int d\mathbf{r} \rho(r) e^{i\mathbf{G} \cdot \mathbf{r}}$ over the spherical volume element $d\mathbf{r} = r^2 dr d\varphi \sin(\theta) d\theta$, with $0 \leq r < R$, $0 \leq \varphi \leq 2\pi$ and $0 \leq \theta \leq \pi$.

Case 1. The surface charge density is first defined

$$\rho(r) = \frac{Q}{4\pi R^2} \delta(r - R) .$$

One thus obtains

$$\begin{aligned} F(G) &= \int d\mathbf{r} \rho(r) e^{i\mathbf{G} \cdot \mathbf{r}} = 2\pi \frac{Q}{4\pi R^2} \int dr r^2 \delta(r - R) \int_0^\pi \sin(\theta) d\theta e^{iGr \cos(\theta)} \\ &= Q \frac{\sin(GR)}{GR} . \end{aligned} \quad (1.113)$$

Case 2. The volume charge density is

$$\rho(r) = \begin{cases} \frac{3Q}{4\pi R^3} & r < R \\ 0 & r > R \end{cases} .$$

One thus obtains

$$F(G) = 2\pi \frac{3Q}{4\pi R^3} \int_0^R dr r^2 \int_0^\pi \sin(\theta) d\theta e^{iGr \cos(\theta)} = Q \frac{\sin(GR)}{GR} . \quad (1.114)$$

It does not make any difference, whether the charge is distributed onto the surface or in the whole sphere!

1.10. Consider two uniformly charged spheres S and S' , with correspondingly different volume charge densities ρ and ρ' and radii R_0 and R'_0 . The two spheres are separated by a distance $a > R_0 + R'_0$. Determine the electric field at a far-away detector, as it results from the two spheres as sources.

Solution. Let z be the axis connecting the centers of the two spheres. Tailoring (1.37) to the present geometrical situation one obtains:

$$E \propto \int_V d\mathbf{r} \rho e^{-i\Delta \mathbf{k} \cdot \mathbf{r}} + e^{-i\Delta \mathbf{k} \cdot \mathbf{a}} \int_{V'} d\mathbf{r} \rho' e^{-i\Delta \mathbf{k} \cdot \mathbf{r}} . \quad (1.115)$$

Using the result of Problem 1.7, (1.115) becomes:

$$E \propto F(qR_0) + e^{iqa \cos(\theta)} F'(qR'_0)$$

in terms of the angle θ that $\Delta \mathbf{k}$ forms with the z -axis.

In the limiting case of two equal spheres, the intensity at the detector would result to be modulated by the factor

$$W = |1 + e^{iqa \cos(\theta)}|^2 = 2[1 + \cos(qa \cos(\theta))] = 4 \cos^2 \left[\frac{qa}{2} \cos(\theta) \right] . \quad (1.116)$$

The quantity W has cylindrical symmetry with respect to the z -axis, so it remains unchanged after the transformation $\theta \rightarrow \pi - \theta$. In particular, one has $W = 4 \cos^2(qa/2)$ for $\theta = 0, \pi$, and $W = 4$ for $\theta = \pi/2$.

Problems without solutions

1.11. Consider a plane and try to cover it by pentagonal tiles. Describe the resulting geometrical characteristics.

1.12. Consider the tetragonal lattices. Calculate the modulus of the reciprocal lattices vectors.

1.13. Consider a simple hexagonal lattice as for graphite. Calculate the intensity of diffracted X-rays for this lattice.

1.14. Consider a square and a simple cubic lattice. Calculate the number of first, second, third, and so on neighbors.

1.15. Draw the Wigner-Seitz cell for a centered rectangular lattice with $a/b = 3$ and $a/b = 1/3$.

1.16. Draw the Brillouin zone for the lattices of Problem 1.14.

1.17. Consider a face-centered cubic lattice with side a and the Miller indexes $(1, 1, 1)$ and $(3, 2, 1)$. Find the distance between the planes within the corresponding families of planes.

1.18. A cylindrical uniform charge distribution of radius R_0 and height $2h$ has its center in the point $(0, 0, 0)$. Calculate the electric field in the plane $z = 0$ at far distances r from the origin $(0, 0, 0)$.

1.19. Consider the Lennard-Jones potential in the form

$$U(R) = \varepsilon \left[\left(\frac{\sigma}{R} \right)^{6n} - 2 \left(\frac{\sigma}{R} \right)^6 \right].$$

Find the equilibrium positions for $n = 0, 1, 2$, and 3 .

1.20. A material characterized by a simple cubic structure with side a , has dimension (Na, Na, Na) . The cohesive energy for a bulk atom is E_0 , the one for a surface atom is $E_0/2$, the one for a side atom is $E_0/4$, and the one for a vertex atom $E_0/8$. Calculate the total energy of the system.

1.21. Two electric dipoles, each composed of two charges $+q$ and $-q$ at distance d , are aligned so that their principal symmetry axes, namely those containing the two charges, coincide. The distance between the dipoles centers is $a \gg d$. Calculate the potential energy of the system in the two cases of parallel and antiparallel dipoles.

1.22. Consider $2N$ dipoles aligned as in Problem 1.21, that is along the perpendicular axis to the dipoles. Calculate the potential energy of the dipole at the origin.

1.23. Consider the two-dimensional free electron gas with density n . Show that the Fermi wavenumber is given by $k_f = \sqrt{2\pi n}$.

References

1. L. Bloomfield: *How Everything Works. Making physics out of ordinary*. John Wiley and Sons, USA (2008)
2. N.W. Ashcroft and N.D. Mermin: *Solid State Physics*. 2nd ed. Holt-Saunders, Tokyo (1976)
3. F. Bassani and U.M. Grassano: *Fisica dello stato solido*. 1st ed. Bollati Boringhieri, Torino (2000)
4. N. Nilius, T.M. Wallis and W.Ho, *Appl. Phys. A* **80**, 951 (2008)
5. N. Nilius, T.M. Wallis and W. Ho, *Science* **297**, 1853 (2002)
6. G.S. Rohrer: *Structure and Bonding in Crystalline Materials*. 1st ed. Cambridge Univ. Press, Cambridge (2004)
7. Guozhong Cao: *Nanostructures and nanomaterials*. 1st ed. Imperial College Press, London (2004)
8. C.R. Kagan, C.B. Murray and M.G. Bawendi, *Phys. Rev. B* **54**, 8633-8643 (1996)
9. B. A. Korgel and D. Fitzmaurice, *Phys. Rev. B* **59**, 14191-14201 (1999)
10. H. Hofmeister, F. Huisken and B. Kohn, *Eur. Phys. Jour. D* **9**, 137-140 (1999)
11. M. D. Klein M.D., G. K. Horton and J. L. Feldman, *Phys. Rev.* **184**, 968-978 (1969)
12. D. N. Batchelder, P. L. Losee and R. O. Simmons, *Phys. Rev.* **162**, 767-775 (1967)
13. E. R. Dobbs and G. O. Jones, *Rep. Prog. Phys.* **20**, 516-564 (1957)

Chapter 2

Electronic structure of nanosystems and crystals

Abstract The electronic state of a system of spatially confined atoms is first studied within a single-particle Hamiltonian approach, starting from atomic orbitals with suited boundary conditions. Then, the case of a macroscopic crystal is obtained after determining the limiting behavior of the confined system while the number of atoms becomes infinitely large. As an alternative to the atomic orbitals, the same electronic states are as well determined starting from a different basis set, that is composed of free-particle wavefunctions. The Bloch theorem is introduced, as the fundamental tool establishing the form assumed by the wavefunctions representing the electronic states, it being a consequence of lattice translational invariance and periodicity. Specific methods suited to the computation of electronic states are thus illustrated, that are applied to selected semiconductors, graphene, and carbon nanotubes. Then, a few elementary concepts of many-particle physics are introduced, useful to describe the interactions of electrons in crystals and nanostructures and to introduce different types of correlated ground states. Indeed, interacting ground states are first investigated within the Hartree and Hartree-Fock methods and excitonic effects are considered. Eventually, ground states related to spin-related magnetic phenomena and to superfluidity and superconductivity are described at an elementary level.

2.1 Introduction

Did you ever ask yourself how xerographic copiers, electric power distribution, generators and motors work? Or else how it is possible to record or store huge amounts of information in magnetic disks? Or how levitating trains work? This content of the present Chapter opens up the concepts useful to answer these and related questions [1].

Chap. 1 has been devoted to the study of the structure of crystals and nanosystems. The Chapter ended with the classification of crystals and of their properties as they result from the nature of different bonding mechanisms. They are microscopically built up from the electron wave functions of individual atoms. The ordered

crystal structure modifies the electronic states of individual atoms into the structure of the electronic energy states of the whole crystal or nanosystem. How this modification occurs and its consequences, is the subject of the present Chapter.

The first part of the Chapter illustrates how the electronic states of nanocrystals can be constructed from those of the atoms arranged in the lattice, starting from the so-called tight-binding method: basic ingredients here are the system Hamiltonian and electron wavefunctions of the atoms composing the nanostructure. The tight-binding method is used first, since it is intuitively simple and especially useful to follow the evolution of energy levels from the single atom up to the realization of a macroscopic crystal obtained after progressive increase of the number of atoms in the system. In such a bottom up approach, the evolution from nanoscopic to macroscopic crystals is exploited to comprehend the scale laws connecting the electronic structure to the crystal size [2].

The second part of the Chapter considers the electronic structure of macroscopic crystals, yet by means of the tight-binding method after suited modifications to evolve from finite to infinite system. Then alternative methods are introduced bearing different advantages and disadvantages, namely Plane-Wave, Wannier and Pseudopotential methods. The methods are applied to nanostructures as well, with special attention to the electronic states of three-dimensional nanostructures with spherical geometry. These methods progressively evolve from a description in terms of localized states given by the atomic wavefunctions to fully delocalized states given by plane waves. Plane waves are periodic functions by construction, this is the main reason why they are especially useful to describe macroscopic crystals [3, 4, 5, 6, 11].

The general problem of studying electronic states starting from electron-electron and electron-nuclei interactions is very complex. Solutions to the problem require suited approximations or powerful computational methods based on first principles. Thus the third part of the Chapter introduces the Hartree and Hartree-Fock methods along with basic concepts of many-particles physics, as elementary approaches to account for electronic interactions. In fact, these methods lead to the possibility of effectively describing the system of interacting particles in terms of a system of independent quasi-particles subjected to an effective appropriately constructed single-particle potential. The validity of Hartree and Hartree-Fock approaches seem here to be limited to systems of many weakly interacting particles. Interestingly, the concept underlying this mapping becomes a quite general concept and powerful tool in Chap. 6: there, it is illustrated how the treatment of even strongly interacting and inhomogeneous systems is possible within the Density Functional Theory framework.

The last part of the Chapter is devoted to introduce two pivotal concepts that open up wide scenarios in many-particle physics, that are subject of current and frontier research. First, the idea that electrons or more generally particles in a crystal have a spin, besides a mass and a charge: interactions of particle spins with external fields or with other spins determine the magnetic properties of materials and may also lead to entirely novel ground states, starting from the best well known ferro- and antiferromagnets. Second, the idea that pairing may occur which entangles single particles

in ways that cannot be traced back to combinations of single-particle electronic orbitals, irrespective of modifications they can undergo because of interactions: pairing in correlated many-particle systems produces entirely novel ground states, starting from superconductors and superfluids. These two types of novel ground states are just examples of a wide variety of fascinating possibilities and related applications, that current theoretical research and experimental realizations bring at hand, for which the interested student is referred to more advanced textbooks [12] and references therein.

2.2 Nanocrystals in one dimension

The simplest possible case is treated first: a nanocrystal in one dimension, that is a chain with a finite number of atoms arranged on a substrate, for engineering and synthesis reasons described in Chap. 1.

Intuition helps realizing that wavefunctions of atoms in a crystal or nanocrystal keep at least part of the characteristics of the individual atoms. It is known that atomic wavefunctions are spatially localized around the nucleus, according to the energy of the state:

Definition

The so-called core states, that are those of the fully occupied shells, possess large binding energies and are tightly localized close to the nuclei.

On the other hand,

Definition

The so-called valence states, that correspond to the partly occupied shells, possess lower binding energies and are more loosely localized.

Fig. 2.1 sketches the qualitative behavior of core (a) versus valence (b) wavefunctions. In contrast with core states, valence wavefunctions are characterized by the presence of a significant component in the region between each two atoms. Remarkably, valence states are also characterized by oscillatory behavior in proximity of individual atoms. These oscillations ensure orthogonality between different states: indeed, wavefunctions are characterized by a vanishing overlap integral.

In both cases, the oscillations are very similar to those of individual isolated atoms. If the distance between atoms were increased, the valence wavefunctions displayed in Fig. 2.1 would indeed progressively evolve towards those of the isolated atoms. Clearly, the process would occur more rapidly for the core states.

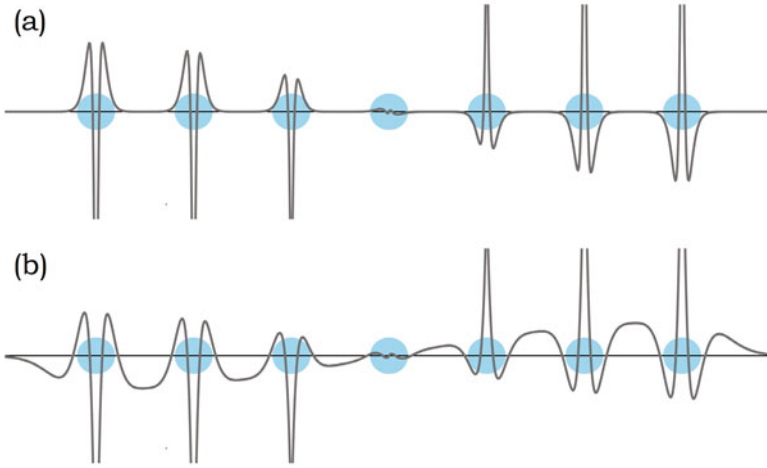


Fig. 2.1 One dimensional nanocrystal. Sketch of the qualitative behavior of core (a) and valence (b) wavefunctions

2.2.1 Tight-Binding method

The considerations discussed in Sec. 2.1 suggest that a suited starting expression representing the states of a nanocrystal be a linear combination of atomic states centered on the lattice sites \mathbf{R}_n that the atoms occupy:

$$\psi(\mathbf{r}) = \sum_{n=1}^N \sum_{\alpha} \phi_{n,\alpha} \varphi_{\alpha}(\mathbf{r} - \mathbf{R}_n). \quad (2.1)$$

In (2.1), $\varphi_{\alpha}(\mathbf{r} - \mathbf{R}_n)$ represent the atomic orbitals and index α labels the full set of the corresponding quantum numbers. The coefficients $\phi_{n,\alpha}$ are to be determined. Use of (2.1) as a starting expression to determine the electronic states corresponds to the Tight-Binding (TB) Method. In general,

Concept

The results provided by the TB method are progressively better approximations to reality, as long as electrons are more strongly bound to the respective atoms, becoming worst in the opposite case of more loosely bound electrons.

The success of the TB method washes away while the situation evolves from Fig. 2.1 (a) towards Fig. 2.1 (b).

For easier equation handling, the Dirac notation is from now on used, so that the orbital $\varphi_{\alpha}(\mathbf{r} - \mathbf{R}_n)$ is labeled by $|\mathbf{R}_n, \alpha\rangle$. Given the single-particle nanostructure Hamiltonian H , the eigenvalue equation determining the eigenstates is

$$H\psi(\mathbf{r}) = E\psi(\mathbf{r}).$$

A set of equations for the unknown coefficients $\phi_{n,\alpha}$ is needed. This is obtained after projecting onto the states $|\mathbf{R}_n, \alpha\rangle$, namely:

$$\sum_{n=1}^N \sum_{\alpha} \langle \mathbf{R}_m, \beta | H | \mathbf{R}_n, \alpha \rangle \phi_{n,\alpha} = E \sum_{n=1}^N \sum_{\alpha} \langle \mathbf{R}_m, \beta | \mathbf{R}_n, \alpha \rangle \phi_{n,\alpha}. \quad (2.2)$$

Notice the presence of overlap integrals in the right-hand side of the equation, showing that this actually is a generalized eigenvalue problem.

The size of the matrices depends on the number of both sites and orbitals. As an example, isolated copper atoms have atomic number 29. The electronic configuration of copper is indeed $[\text{Ar}]3d^{10}4s^1$, leading to a total of 15 orbitals. The electrons completely fill in the shells with $n = 1, 2$, and 3 , while the shell with $n = 4$ is partly occupied, in particular by one single electron. Thus, if N is the total number of copper atoms in the nanocrystal, the corresponding matrices are as large as $15N \times 15N$. The matrix representation significantly simplifies when the different degree of localization of core and valence atomic orbitals is taken into account: the cross matrix elements between core and valence states are vanishingly small, because the corresponding overlap integrals have negligible values. Following the above reasoning, the matrix can approximately be reduced to two main independent blocks, as schematically shown in Fig. 2.2. In practice, core and valence solutions can be treated in independent calculations. This is clearly an approximation, that nicely works in most cases, though.

More complicated situations can be addressed by refining this approximation. In particular, core and valence blocks can be usefully redefined, with selected core elements moved into the valence part. The example of copper is once more of help. As for other transition elements indeed, d-states of copper are characterized by an intermediate degree of localization between core-like and valence-like states. A more accurate computation of the electronic structure would result from moving the d core states into the valence block. The matrix structure of Fig. 2.2 is amenable to treat also the case of excited states. In this case the valence block can be extended to include the first higher and unoccupied atomic sub-shells.

On general grounds, electronic states $\psi(\mathbf{r})$ can be demonstrated to be exactly represented by an expression similar to (2.1), where however the functions associated and centered on the lattice sites are to be different from atomic wavefunctions of isolated atoms. These new functions, named after Wannier and from now on la-

Fig. 2.2 Tight-Binding method. Schematic block matrix representation of the equation set for the unknown coefficients given by (2.2)

$$\begin{pmatrix} \text{CORE} & 0 \\ 0 & \text{VALENCE} \end{pmatrix}$$

beled by $w_p(\mathbf{r} - \mathbf{R}_n)$, can be shown to be to be rigorously orthogonal to each other, that is

$$\int d\mathbf{r} w_p^*(\mathbf{r} - \mathbf{R}_n) w_{p'}(\mathbf{r} - \mathbf{R}_m) = \delta_{p,p'} \delta_{n,m} . \quad (2.3)$$

The Wannier functions $w_p(\mathbf{r} - \mathbf{R}_n)$ thus acquire a significant meaning. They are explicitly constructed in Sec. 2.4.2, for the easier case of an infinite crystal.

Reverting back to the case of the model nanostructure under consideration, all possible complications described above are now neglected focusing the attention on the valence block built up solely from the s-orbitals. Equation (2.2) can be written as

$$\sum_{n=1}^N [\langle \mathbf{R}_m | H | \mathbf{R}_n \rangle - E \langle \mathbf{R}_m | \mathbf{R}_n \rangle] \phi_n = 0, \quad (2.4)$$

where the no longer useful index α has been eliminated. An additional simplification to the equation set (2.4) occurs whenever the overlap integrals $\langle \mathbf{R}_m | \mathbf{R}_n \rangle$ can be neglected, precisely setting $\langle \mathbf{R}_m | \mathbf{R}_n \rangle = \delta_{m,n}$. This is the case when the atomic orbitals can be well approximated by Wannier functions. Equation (2.4) thus simplifies into:

$$\sum_{n=1}^N (\langle \mathbf{R}_m | H | \mathbf{R}_n \rangle - E \delta_{m,n}) \phi_n = 0. \quad (2.5)$$

2.2.2 Matrix elements

The matrix elements $\langle \mathbf{R}_m | H | \mathbf{R}_n \rangle$ are now to be calculated. To this aim, an explicit form of the Hamiltonian is required. Within the adopted scheme of independent particles, the Hamiltonian operator acquires the single-particle form

$$H = -\frac{\hbar^2}{2m} \nabla^2 + \sum_n v(\mathbf{r} - \mathbf{R}_n) . \quad (2.6)$$

The first term represents the kinetic energy, while the second is the potential energy composed by the sum of potential-energy terms centered on the single atoms.

The summation over the lattice sites appearing in (2.6) suggests to classify the matrix elements $\langle \mathbf{R}_m | H | \mathbf{R}_n \rangle$ according to the number of centers that are involved. For example, the diagonal matrix element $\langle \mathbf{R}_m | H | \mathbf{R}_m \rangle$ involves integrals of functions centered both on the same site $\langle \mathbf{R}_m | v(\mathbf{r} - \mathbf{R}_m) | \mathbf{R}_m \rangle$ and on different sites $\langle \mathbf{R}_m | v(\mathbf{r} - \mathbf{R}_n) | \mathbf{R}_m \rangle$, with $n \neq m$. The matrix elements $\langle \mathbf{R}_m | H | \mathbf{R}_n \rangle$, with $n \neq m$, add to the above type of one- and two-centers integrals also integrals with three centers. This is sketched in Fig. 2.3: the overall degree of overlap of functions centered on three different centers is expected to be tinier than in the case of two and one centers. In most cases, three-centers integrals can be neglected with respect to one- and two-center integrals. Clearly, the validity of this approximation has to be estimated for each single situation.

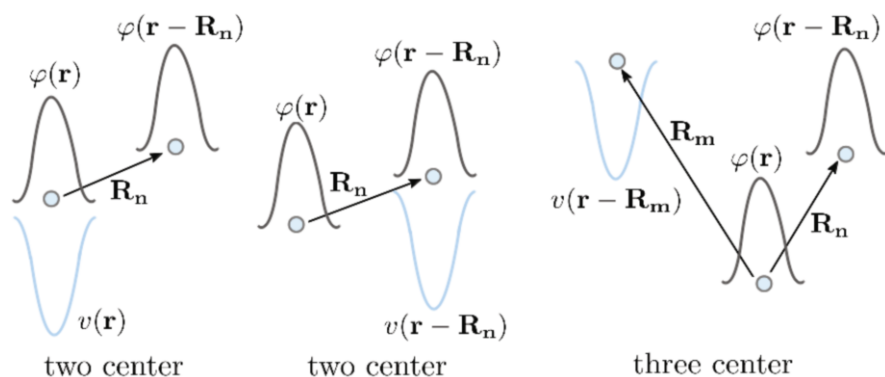


Fig. 2.3 Determination of the matrix elements in TB method. The meaning of integrals involving two and three centers is schematically depicted. The general case of a 3D spatial arrangement of atomic sites is represented. The vectors connecting selected pairs of sites are shown in the figure

The calculation of the matrix elements is the most committing step in the determination of electronic states in a nanocrystal. Students should be aware of this aspect. The atomic wavefunctions are often known only in numerical form, that is as tables of numbers, and the check of accuracy of the integrals might become an additional issue to be taken care of. An alternative strategy exists, which considers the matrix elements as parameters to be determined after fitting appropriate sets of experimental data, an idea introduced already back in the mid 20th century by Slater [6]. This approach is known as Semi-empirical Tight-Binding and has been demonstrated to be very efficient in studies of electronic states in both crystalline and, more recently, nanostructure materials. In essence, once the non-vanishing (non-negligible) matrix elements have been singled out, treating them as empirical parameters has the meaning of feeding in them all the unknown effects that have been forcedly set to zero. This is partly the reason why best results are obtained in spite of neglecting the overlap integrals. The Semi-empirical TB is later applied to selected examples.

2.2.3 Electronic structure

The Slater method is illustrated through the following case, where the fitted parameters play an important role even if they are not explicitly calculated. Here is the schematic procedure.

Procedure

Step 1. Begin with (2.4), or with (2.5) if the overlap integrals can be neglected (see Sec. 2.2.4).

Step 2. Compute the matrix elements $\langle \mathbf{R}_m | H | \mathbf{R}_n \rangle$.

Step 3. Choose the boundary conditions.

Step 4. Solve the eigenvalue problem with respect to E and ϕ_n , paying attention to the allowed values of ϕ_n that are dictated by the boundary conditions.

Step 5. Compose the nanocrystal wavefunction according to (2.1).

Let the procedure be tested through an example.

Examples

In order to make the essence of the procedure as much clear as possible, an additional simplification is introduced in which the matrix elements between non first neighbor atoms are neglected. One has then:

$$\langle \mathbf{R}_m | H | \mathbf{R}_n \rangle = \begin{cases} E_0 & \text{if } m = n, \\ \gamma(\delta_{m,n+1} + \delta_{m,n-1}) & \text{if } m \neq n. \end{cases} \quad (2.7)$$

E_0 and γ are the numerical values of the matrix elements, which from now on are considered as parameters. Expression (2.7) does not include the cases where index m or n , or both, label sites 1 and N at either one end of the chain. These are special atoms, that are in contact with the external environment. Thus on physical grounds one could expect that E_0 and γ might assume different numerical values for the boundary atoms with respect to the bulk ones. This issue is better taken care of in Sec. 2.2.5, and neglected at this stage. Inserting the matrix elements (2.7) into (2.5), the equation set has the form:

$$E_0 \phi_n + \gamma(\phi_{n+1} + \phi_{n-1}) = E \phi_n \quad \text{with } n = 1, \dots, N. \quad (2.8)$$

The set of equations (2.8) requires its own boundary conditions to be set. Different physical situations are amenable to different boundary conditions, as detailed in Chap. 3. One among the simplest are the so-called hard-wall boundary conditions

$$\phi_0 = 0, \quad ; \quad \phi_{N+1} = 0, \quad (2.9)$$

where sites with $n = 0$ and $n = N + 1$ are taken to be empty. If

$$\phi_n = A e^{\pm i k n a} \quad (2.10)$$

were to be considered, k labeling the wavenumber, it is immediately seen by inspection that ϕ_n is a solution of system (2.8) with energy

$$E = E_0 + 2\gamma \cos(ka). \quad (2.11)$$

However, (2.10) does not satisfy the boundary conditions (2.9). One could then try a linear combination of e^{ikna} and e^{-ikna}

$$\phi_n = A \sin(kna) , \quad (2.12)$$

and check that this leads to the same energy given by (2.11) and at the same time that it satisfies (2.9) whenever

$$k = \frac{v\pi}{a(N+1)}, \text{ with } v = 1, 2, 3, \dots, N . \quad (2.13)$$

The constant A has to be determined from the normalization condition for the wave-function. Carrying on the calculation, leads to the results

$$\psi_v(\mathbf{r}) = \sqrt{\frac{2}{N+1}} \sum_{n=1}^N \sin\left(\frac{v\pi}{N+1}n\right) |\mathbf{R}_n\rangle , \text{ with } v = 1, 2, 3, \dots, N . \quad (2.14)$$

Expression (2.11) along with (2.13) represents the energy of level n . These are characterized by the on site energy E_0 coming from the diagonal elements, and by the first-neighbor interaction energy γ . It should be remarked that

Concept

E_0 is generally different from the energy of the isolated atom, since it effectively includes the contribution of the potential energy of the other atoms. This contribution is generally referred to as the crystal field term.

The parameter γ contains significant information as well, which are now being enlightened. First, let the energy difference between the lowest state with $v = 1$ and the highest state with $v = N$ be computed. Assuming for example $\gamma < 0$ one has

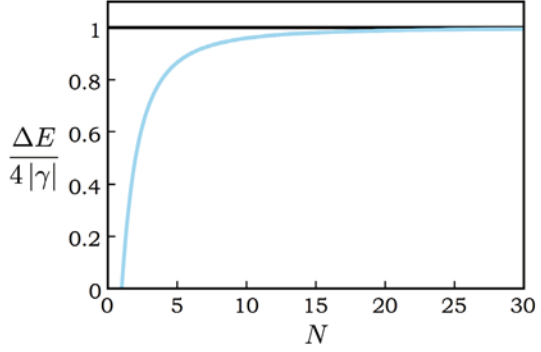
$$\Delta E = E_N - E_1 = -2|\gamma| \left[\cos\left(\frac{N\pi}{N+1}\right) - \cos\left(\frac{\pi}{N+1}\right) \right] . \quad (2.15)$$

In the limit $N \rightarrow \infty$, $\cos[N\pi/(N+1)] \approx \cos[\pi - \pi/(N+1)]$, and (2.15) can be expanded in $1/N$ power series. Expansion leads to

$$\frac{\Delta E}{4|\gamma|} = 1 - \frac{\pi^2}{N^2} + \dots . \quad (2.16)$$

Expression (2.15) along with (2.16) thus states that while the number N of atoms is increased, i.e. the nanocrystal is evolving towards a bulk crystal, the energy levels of the nanocrystal arrange themselves into an energy range, whose value progressively shrinks to a constant, precisely to $4|\gamma|$. The convergence is rapid according to the N^{-2} term on the right-hand-side of (2.16). This behavior is sketched in Fig. 2.4.

Fig. 2.4 Electronic structure of a 1D nanocrystal. Evolution of the energy range $\Delta E/4|\gamma|$ in (2.15), as a function of the number of atoms N , from the nanocrystal case with $N \rightarrow 1$ towards the bulk crystal case with $N \rightarrow \infty$



Second, look at the energy spacing $\Delta E_v = E_{v+1} - E_v$ between adjacent levels. For finite v and $N \rightarrow \infty$, one has $\Delta E_v = 2|\gamma|\pi^2(v + 1/2)/(N + 1)^2$, yet scaling as N^{-2} . Thus, besides the shrinking of the energy range, with increasing N the energy levels tend to thicken into an energy band as wide as $4|\gamma|$. Along with the energy levels, the allowed values of k in (2.13) become denser. The range of k vectors into which this thickening occurs is the $N \rightarrow \infty$ limit of (2.13), that is $(0, \pi/a)$, obtained while the spacing between adjacent k vectors shrinks to zero as N^{-1} . Problem 2.14 discusses a simple case.

Third, consideration of the electron spin brings additional information. Due to the spin indeed, each state of the nanostructure can be occupied by two electrons. If the element composing the nanostructure is monovalent, only one half of the states is therefore occupied. At zero temperature, that is the case under study, the energy of the highest occupied state is defined to be the Fermi energy or Fermi level. It is usefully reminded here from Appendix 1.10 of Chap. 1 for a three-dimensional system, that given the number N of free electrons and therefore their density n , the latter dictates the Fermi wavevector k_f

$$k_f = (3\pi^2 n)^{\frac{1}{3}}, \quad (2.17)$$

as the wavevector corresponding to the last occupied state with the Fermi energy

$$\varepsilon_f = \frac{\hbar^2}{2m} k_f^2 = \frac{\hbar^2}{2m} (3\pi^2 n)^{\frac{2}{3}}. \quad (2.18)$$

Here, ε_f is the maximum possible energy at zero temperature. While N gets to be very large, the difference in energy between the Fermi level and the first unoccupied state is vanishingly small, a circumstance that identifies a good conductor of electric current. The connection between the availability of states in proximity of the Fermi level and electric conductivity of materials is studied in Chap. 4.

Overall, from the use of Semi-empirical Tight Binding method one may conclude that

Concept

While the nanocrystal evolves towards a bulk crystal with $N \rightarrow \infty$, three relevant facts occur. The energy levels thicken as N^{-2} inside an energy range that progressively shrinks to a constant value, eventually getting for the bulk the band-width value $4|\gamma|$ dictated by the first neighbor matrix element γ . Along the same lines, the wavenumbers k thicken as N^{-1} inside a range, that tends to become $(0, \pi/a)$. Finally, unoccupied states become available at an energy cost that tends to be vanishingly small and thus to favor electric conduction processes.

The wavefunctions (2.14) resulting in the case of a chain with four atoms are displayed in Fig. 2.5. Dashed lines represent the coefficients $\sin[\nu\pi n/(N+1)]$, the variable n being here treated as a real number. Remarkably, these functions are the envelop to those in (2.14), a remark that will be very useful later on.

Expression (2.8) suggests a very simple representation for the energy of an electronic state. Let the coefficients ϕ_n be a very slowly-varying function of n . In this case, a continuum $x = na$ can be usefully introduced, leading to the expression

$$E_0\phi(x) + \gamma[\phi(x+a) + \phi(x-a)] = E\phi(x). \quad (2.19)$$

To the second order in Taylor expansion of $\phi(x+a)$ with respect to a around $a = 0$, one has:

$$\begin{aligned} \phi(x+a) &\simeq \phi(x) + a\frac{d\phi}{dx} + \frac{1}{2}a^2\frac{d^2\phi}{dx^2}, \\ \phi(x-a) &\simeq \phi(x) - a\frac{d\phi}{dx} + \frac{1}{2}a^2\frac{d^2\phi}{dx^2}, \\ \phi(x+a) + \phi(x-a) &\simeq 2\phi(x) + a^2\frac{d^2\phi}{dx^2}. \end{aligned} \quad (2.20)$$

Thus, (2.19) becomes

$$(E_0 - E + 2\gamma)\phi(x) + \gamma a^2\frac{d^2\phi}{dx^2} = 0, \quad (2.21)$$

that resembles a Schrödinger equation for a single free particle with energy

$$E = E_0 + 2\gamma + \frac{\hbar^2}{2m^*}k^2, \quad (2.22)$$

and mass

Definition

$$m^* = -\frac{\hbar^2}{2\gamma a^2}, \quad (2.23)$$

that defines the so-called effective mass.

The effective mass corresponds to a significant and useful concept:

Concept

Independent particles interacting with the crystal potential can be represented as if they were free particles with an effective mass. Such effective particles are referred to as quasiparticles.

The effective mass can be alternatively obtained starting from expression (2.11) for the eigenvalues and expanding $\cos(ka)$ up to second order in ka around $ka = 0$. In this case one would obtain $E = E_0 + 2\gamma - \gamma(ka)^2$, that coincides with the expression for the eigenvalues of (2.21), once the definition of effective mass is taken into account.

The concepts and procedures illustrated in the present section correspond to the minimal knowledge useful to understand and predict the electronic structure of nanocrystals, embodying also the main concepts related to bulk crystals. The next subsections are devoted to introduce one complication at a time, yet remaining within this simple and pedagogical, though realistic, chain model. First a non-negligible overlap integral, then the effect of surface atoms, namely those lying at the boundaries, and finally the effects resulting from having more than one species per cell. Applications to real nanocrystals are eventually discussed, before preparing to step into the electronic structure of an infinite bulk crystal.

Quick Questions

Q1. Provide a physically sound and qualitative explanation for the thickening of electronic states into an energy band of finite width [Hint: consider the effect of interactions].

Answer. Interactions between a given electron and an atom at large enough distance become negligible. Thus, adding further atoms to a system with size larger than a critical value does not modify the properties of the states.

Q2. The mass of a particle is intrinsically represented by a positive number. Does this statement hold for the effective mass of a quasiparticle too?

Answer. No: the sign of the effective mass depends on the sign and size of γ , that is not determined a priori.

2.2.4 Tight Binding with overlap

The assumption $\langle \mathbf{R}_m | \mathbf{R}_n \rangle = \delta_{m,n}$ of the previous section is here being removed, that is the overlap integrals are no longer negligible. The results of Sec. 2.2.3 are unchanged in their essence, a few new details arising from this added complication. Consider to be in the condition $\langle \mathbf{R}_m | \mathbf{R}_n \rangle \neq \delta_{m,n}$. Expression (2.8) is easily transformed into:

$$E_0 \phi_n + \gamma(\phi_{n+1} + \phi_{n-1}) = E [\phi_n + s(\phi_{n+1} + \phi_{n-1})] , \quad (2.24)$$

where now $\langle \mathbf{R}_m | \mathbf{R}_{m\pm 1} \rangle = s$. Closely following the same considerations, ϕ_n in (2.12) still results to be solution of (2.24). The eigenvalues instead become

$$E = \frac{E_0 + 2\gamma \cos ka}{1 + 2s \cos ka} , \quad (2.25)$$

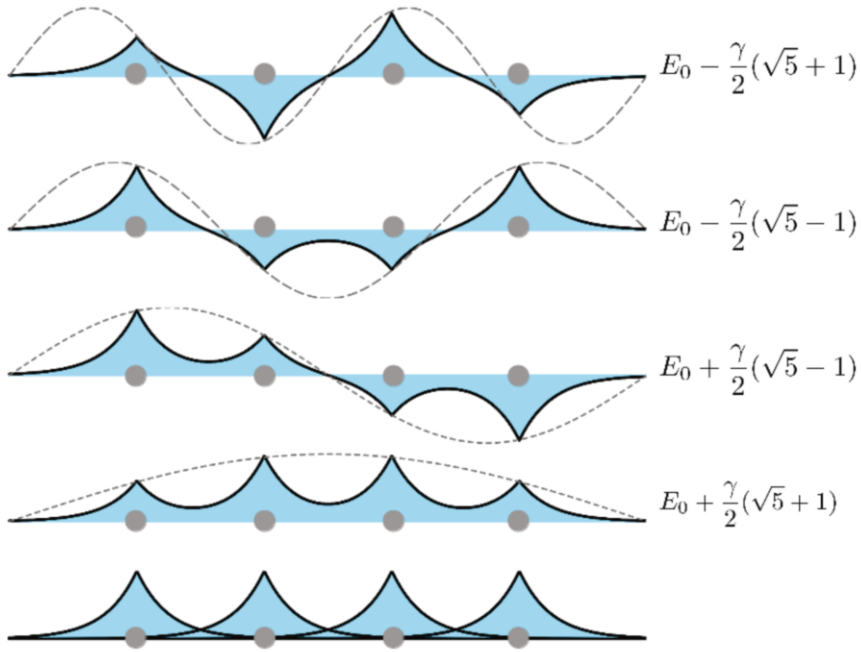


Fig. 2.5 Electronic structure of 1D nanostructures. Chain with four atoms, the grey dots highlighting the atomic positions. Top four panels: the wavefunctions of the first four electronic states from (2.14) as represented by solid lines. The resulting eigenvalues are reported in each case. The coefficients $\sin(\nu\pi n/(N+1))$ are displayed as dashed lines, after treating n as a continuum variable: note that they envelope the wavefunctions as represented by (2.14). Bottom panel. The starting atomic wavefunctions are displayed

where the allowed k values are once more determined by (2.13), the boundary conditions being unchanged. As for the (2.14), the new eigenfunctions are now

$$\psi_v(\mathbf{r}) = \sqrt{\frac{2}{(N+1)(1+2s\cos\frac{v\pi}{N+1})}} \sum_{n=1}^N \sin\left(\frac{v\pi}{N+1}n\right) |\mathbf{R}_n\rangle. \quad (2.26)$$

Thus, it is obtained that

Concept

Finite values of the overlap integrals between states centered on adjacent sites modify the energies of the electronic states and leave the wavefunctions unchanged.

A criterion would be useful to help establishing whether these overlap integrals should be better included or not. As a thumb rule, if the TB method is used by direct computation of the matrix elements, the states overlap cannot be neglected. If the matrix elements are instead used as parameters to be fitted from the experimental data as in Slater's method, neglecting the overlap could be a convenient move. It is clear that the calculation of matrix elements by means of these two approaches can produce quite significantly different results.

2.2.5 Surface and impurities states

Surface states. The additional effect is now investigated, that atoms at the boundaries of the chain might in general have different values of parameters E_0 and γ . These first and last atoms at the chain boundaries are the simplest example of surface atoms. The inclusion of such a complication represents in fact practical ways of approaching surface effects, starting from a simple situation.

Both the on site and first-neighbors matrix elements coupling the surface atoms to the immediately adjacent internal atoms, must acquire values different from the others. To make life simpler, assume that the only the on site matrix element of the surface atom at site $n = 1$ be different from all the others, its value amounting to E'_0 . The set of equations (2.8) remains unchanged, except the equation line corresponding to $n = 1$. This now becomes

$$(E'_0 - E) \phi_1 + \gamma \phi_2 = 0. \quad (2.27)$$

In essence, a new boundary condition substitutes that represented by the first between the two equations in (2.9). The boundary condition $\phi_{N+1} = 0$ is instead left unchanged: the issue here is to estimate the qualitative effect of the surface atom at position $n = 1$ on the properties of the states, rather than the detailed problem. This time, the function given by (2.12) does not satisfy the new boundary condition

(2.27). Following a reasoning analogous to that leading to (2.12), let ϕ_n be this time a linear combination of $\sin(kna)$ and $\cos(kna)$. That is:

$$\phi_n = A \sin(kna + \delta) . \quad (2.28)$$

By inspection, expression (2.28) now satisfies (2.27). As demonstrated in Problem 2.1, the corresponding energies are yet given by (2.11), provided that

$$\tan \delta = - \frac{(E'_0 - E_0) \sin(ka)}{(E'_0 - E_0) \cos(ka) - \gamma} . \quad (2.29)$$

It is immediately seen that in the limiting case $\delta \rightarrow 0$, the result $E'_0 = E_0$ of Sec. 2.2.4 is recovered.

Expression (2.29) contains two unknowns, k and δ , so that a second equation is to be provided in order to determine the solution. This equation is obtained after imposing that (2.28) satisfies the second of the two expressions in (2.9), namely,

$$ka = \frac{v\pi - \delta}{N+1} , \quad (2.30)$$

with v an integer. Isolating δ from (2.30) and substituting it into (2.29), one obtains

$$\tan [(N+1)ka] = \frac{\sin(ka)}{\cos(ka) - \frac{\gamma}{E'_0 - E_0}} , \quad (2.31)$$

that is an implicit relation for k . If $|\gamma/(E'_0 - E_0)| \geq 1$, (2.31) admits N real solutions for ka . On the other hand, if $|\gamma/(E'_0 - E_0)| < 1$ such equation admits $N - 1$ real solutions. This is shown in Problem 2.10. The missing state is the surface state, i.e. the state localized around the site $N = 1$. Therefore, it can be shown that

Properties

If $|\gamma/(E'_0 - E_0)| \geq 1$, solutions in the interval $0 < ka < \pi$ are delocalized over the whole chain. If $|\gamma/(E'_0 - E_0)| < 1$, one solution appears that is localized around the first edge atom.

The physical nature of the localized surface state is now discussed in a specific example. To this aim, a solution to (2.8) is to be looked for, excepted those with $n = 1$ and $n = N$. Begin with considering the coefficients ϕ_n of the form

$$\phi_n = A e^{ikna} , \quad (2.32)$$

with complex k . After substituting into (2.27), one has

$$0 = E'_0 - E + \gamma e^{[i(x+iy)]} = E'_0 - E + \gamma e^{-y} (\cos x + i \sin x) , \quad (2.33)$$

with $x = \text{Re}(ka)$ and $y = \text{Im}(ka)$. This is meaningful if and only if $x = p\pi$, with p an integer. Thus:

$$e^{-y} = -(-1)^p \frac{E'_0 - E}{\gamma}. \quad (2.34)$$

Considering (2.8) for $n = 2, \dots, N-1$ together with (2.34), the energy E_S of the localized state is found to be

$$E_S = E'_0 + \frac{\gamma^2}{E'_0 - E_0}. \quad (2.35)$$

When substituted into (2.34), it yields

$$e^{-y} = (-1)^p \frac{\gamma}{E'_0 - E_0}.$$

The integer number p has to be chosen so that the right-hand side be positive. A state localized at site $n = 1$ implies the condition

$$\xi \equiv \left| \frac{\gamma}{E'_0 - E_0} \right| < 1.$$

The parameter ξ indeed measures the difference between the surface and bulk on-site matrix elements E'_0 and E_0 with respect to the first-neighbor overlap integral γ .

It is now demonstrated that the energy of the localized surface state is larger than the maximum energy of the band or smaller than the minimum one. The former case is displayed in Fig. 2.6. The minimum energy in the band is $E_m = E_0 - 2|\gamma|$, whereas the maximum is $E_M = E_0 + 2|\gamma|$. Use of (2.35) leads to the conclusion that $E'_0 > E_0$ implies $E_S > E_M$, while $E'_0 < E_0$ implies $E_S < E_m$. This is the subject of Problem 2.9.

The above results allow to express the coefficient ϕ_n in the alternative form

$$\phi_n = A (-1)^n \xi^n, \quad (2.36)$$

showing that the coefficients moduli are decreasing functions of n . The ratio ξ controls the penetration of the wavefunction inside the material: the smaller the ratio, the faster is the rapidity which the wavefunction decays with.

In conclusion,

Concept

Whenever the first atom of the chain has an on-site matrix element different from the other on-site elements inside the chain, the formation of a surface state is possible if $\xi < 1$. Its energy E_S is (2.35) and the coefficients of the wavefunction (2.1) are (2.36). Whenever $\xi \geq 1$, surface states do not occur.

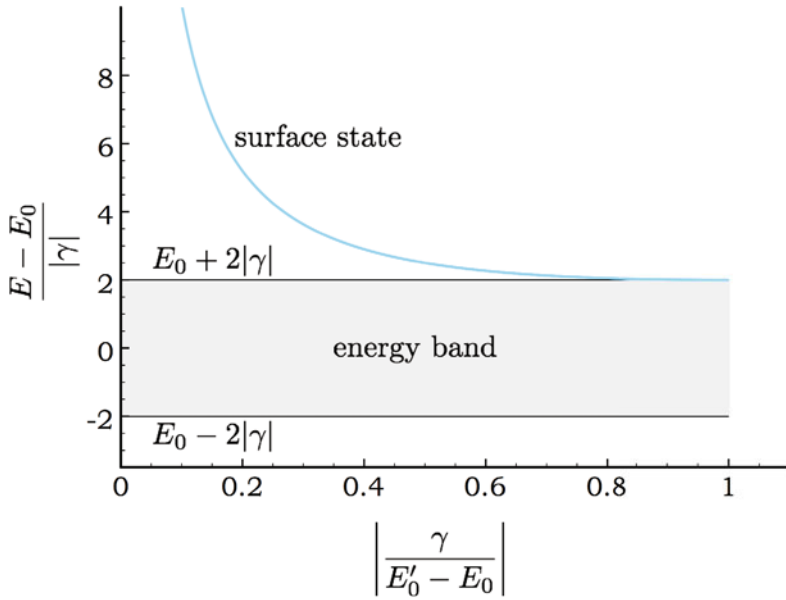


Fig. 2.6 Surface states. The position of the surface-state energy with respect to the energy band

Similar results hold in more complex cases involving macroscopic crystals. Surface states might appear even at the interface between two different materials.

A final remark is in order. The surface state has been obtained by means of (2.32) with a complex k , though it is evident that the exponential cannot fulfill the boundary conditions on site $N + 1$. In fact, (2.32) is the result of an approximation that works nicely when the chain is long enough for the effect of the boundary condition on the $N + 1$ site be negligible. In particular, ϕ_n in (2.32) is anyway vanishing small while the chain length increases, and the boundary condition is satisfied in the limiting sense. The validity of this approximation spoils while the chain shortens. Fig. 2.7 for example shows the energy of a surface state as a function of the number of atoms in the chain, for $\xi = 0.8$. It is seen that while the number of atoms in the chain increases, the exact result converges towards the approximated value provided by the analytical expression (2.35), and represented by the solid line. Agreement between exact and approximated calculations is obtained already for a chain with about 10 atoms. The convergence rapidly depends on how much different are the surface and the bulk on-site matrix elements E_0 , on the scale of the first-neighbor overlap integral γ . Fig. 2.8 completes the information displaying the coefficient ϕ_n of the surface state as a function of the site index from one end to the opposite, as they result from exact diagonalization. The number of atoms in the chain is fixed to be 20, so that convergence to the exact results is reached.

Impurity states. The very same model and procedure can be used to determine the wave function of an impurity state, as it may occur if an atom of different species is

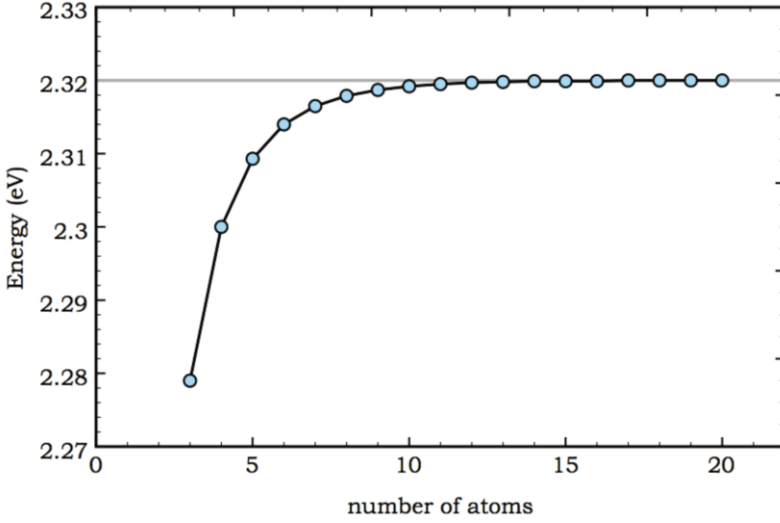


Fig. 2.7 Surface states. The energy of the surface state as a function of the number of atoms in the chain. The results of an exact diagonalization of the Hamiltonian by numerical methods (symbols) are compared with the approximated values from the analytical expression (2.35) (straight line). The calculation parameters are $E'_0 = 2.0$ eV, $E_0 = 1.5$ eV, and $\gamma = -0.4$ eV, so that $\xi = 0.8$

substituted to an atom of the chain. In this case, the on-site energy E_0 of the K -th site is appropriately changed into a different value E'_0 , and then the solution to system (2.8) is looked for. The coefficients ϕ_n are to be appropriately modified, imposing:

$$\phi_n = \begin{cases} A e^{[i(K-n)(x+iy)]}, & \text{for } n \leq K \\ A e^{[i(n-K)(x+iy)]}, & \text{for } n \geq K \end{cases}.$$

Substitution of the above expressions into system (2.8) for $n \neq 1, K, N$, one still obtains $x = p\pi$, with p an integer, and

$$\cosh(y) = (-1)^p \frac{E - E_0}{2\gamma}. \quad (2.37)$$

Substitution into the K -th equation yields

$$e^{-y} = (-1)^p \frac{E - E'_0}{2\gamma}. \quad (2.38)$$

Finally, this relation can be inserted into (2.37), so to obtain the searched energies

$$E_{\pm} = E_0 \pm \sqrt{(E_0 - E'_0)^2 + 4\gamma^2}. \quad (2.39)$$

It is evident that $E_+ > E_M$ and $E_- < E_m$, with E_M and E_m defined above. Considering again $\gamma < 0$, (2.37) and (2.38) admit solutions with real and positive y for the

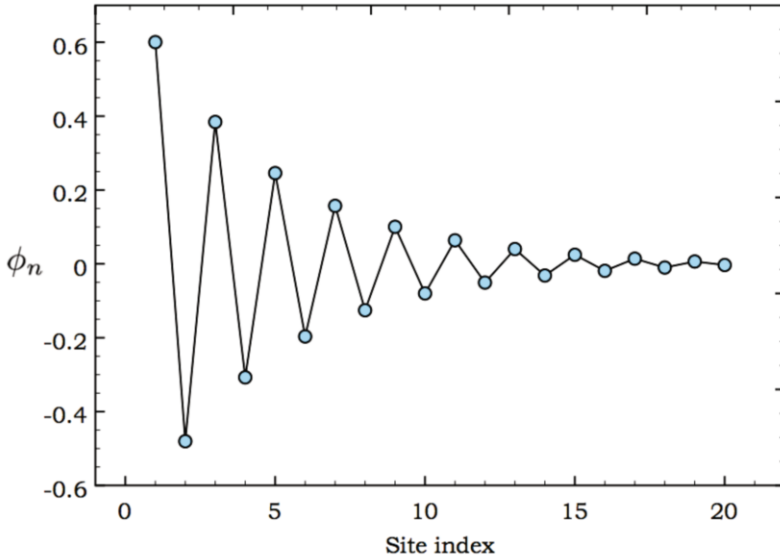


Fig. 2.8 Same as in Fig. 2.7, with the difference that here the coefficient ϕ_n of the surface state is displayed as a function of the site index, running from one end ($n = 1$) of the chain to the opposite ($n = 20$). Symbols are as in Fig. 2.7, referring in both cases to the results of the exact diagonalization for a chain with 20 atoms

E_+ branch if $E'_0 > E_0$ and if $E'_0 < E_0$ for E_- . Once more, these correspond to approximated solutions, suited to the limiting case of a long chain with the K -th site well within it, since the boundary conditions $\phi_0 = \phi_{N+1} = 0$ are not fulfilled. It is usefully remarked that

Concept

The conditions for the existence of impurity states are the same as those for the surface states.

On general grounds, a potential with $E'_0 < E_0$ represents a state possessing an attractive impurity potential, with its energy level taken down from the minimum of the energy band of the electronic states, (2.11). A potential with $E'_0 > E_0$ instead, represents a state possessing a repulsive impurity potential, with its energy level lifted above the maximum of (2.11).

Quick Questions

Q3. Does a surface state exist at the $n = 1$ site, if the amount of first-neighbor overlap γ is modified with respect to the bulk value, instead than E_0 ?

Answer. Yes, since the origin for the existence of a surface state is in any case traced back to a non-uniform behavior of atoms in the chain.

Q4. Does an impurity exist at the $n = K$ site, if γ is modified with respect to the bulk value instead than E_0 ?

Answer. Yes, for the same reason as in Q3.

2.2.6 Chain with $2N$ atoms of two different species

One last complication is here investigated, the consequence of two different species on the behavior of the energy gap between two bands as a function of the cell number N . To this aim, expression (2.1) has to be generalized to take into account the presence of two atomic species. The following convention is adopted: the coefficients ϕ_n with odd (even) n are those referring to the orbitals of the first (second) atomic species, respectively. Considering the interactions between the first neighbors described by parameter γ , the simplest generalization of expressions (2.8) is given by

$$\begin{aligned} (E_0 - E)\phi_1 + \gamma\phi_2 &= 0 \text{ for } n = 1 \\ (E_0 - E)\phi_{2n+1} + \gamma(\phi_{2n+2} + \phi_{2n}) &= 0 \text{ for } n = 1, 2, \dots, N-1 \\ (E'_0 - E)\phi_{2n} + \gamma(\phi_{2n-1} + \phi_{2n+1}) &= 0 \text{ for } n = 1, 2, \dots, N-1 \\ (E'_0 - E)\phi_{2N} + \gamma\phi_{2N-1} &= 0 \text{ for } n = N, \end{aligned} \quad (2.40)$$

with boundary conditions $\phi_0 = \phi_{2N+1} = 0$. Posing

$$\begin{aligned} \phi_{2n+1} &= A \sin[(2n+1)\varphi], \\ \phi_{2n} &= B \sin[2n\varphi] \end{aligned} \quad (2.41)$$

and substituting into the second and third equations in (2.40), one finds that the equations are satisfied if

$$\begin{aligned} (E_0 - E)A + 2\gamma B \cos \varphi &= 0, \\ 2\gamma A \cos \varphi + (E'_0 - E)B &= 0. \end{aligned} \quad (2.42)$$

The first and fourth equations (2.40) are verified if

$$\varphi = \varphi_v = \frac{v}{2N+1} \pi, \quad (2.43)$$

with v an integer assuming the values $v = 1, 2, \dots, N$. From (2.42) the eigenvalues can be calculated, yielding:

$$E_{\pm}(\nu) = \frac{1}{2} \left[E_0 + E'_0 \pm \sqrt{(E_0 - E'_0)^2 + 16\gamma^2 \cos^2 \phi_\nu} \right]. \quad (2.44)$$

Two energy bands show up: $E_-(\nu)$ being an increasing function of ν with a minimum at $\nu = 1$ and a maximum at $\nu = N$; $E_+(\nu)$ being a decreasing function of ν instead, with a maximum at $\nu = 1$ and a minimum at $\nu = N$. The quantity

$$E_+(N) - E_-(N) = \sqrt{(E_0 - E'_0)^2 + 16\gamma^2 \cos^2 \left(\frac{N}{2N+1} \pi \right)} \quad (2.45)$$

provides the size of the gap between the two bands, that is the difference between the minimum value of the highest and the maximum value of the lowest band. The gap size is seen to decrease with increasing N . In a similar way,

$$E_+(1) - E_-(1) = \sqrt{(E_0 - E'_0)^2 + 16\gamma^2 \cos^2 \left(\frac{1}{2N+1} \pi \right)} \quad (2.46)$$

provides the total width of the bands, including their gap. The bandwidth results to be an increasing function of N . The width of the band Γ_1 and Γ_2 are respectively

$$\Gamma_1 = E_-(N) - E_-(1),$$

$$\Gamma_2 = E_+(1) - E_+(N).$$

Notice that these quantities increase when N increases.

Concept

The presence of different atomic species in the same chain modifies the electronic structure with the appearance of additional bands, with given width and separated by a forbidden energy gap. The larger is the number N of cells, the smaller is the gap size and the larger the bandwidths. Both the gap size and the bandwidth approach a constant in the limit of infinite crystal.

2.2.7 Applications to chains of metallic atoms

The concepts developed up to now can be illustrated by means of experimental data on the energy states of chains of metallic atoms. Chap. 1 has anticipated one among possible ways to build up an atomic chain on a substrate, that is using an STM. STM has also been discussed as a fundamental tool to visualize the crystal structure of surfaces, through the measurement of the tunneling current between the probe tip and the surface atoms. Once the nanostructure is realized, this can be characterized by a measurement of the differential conductance dI/dV while the tip is scanned

along the chain and the applied bias is varied. This type of measurement is known as Scanning Tunneling Spectroscopy or STS. The differential conductance dI/dV depends on the number of states per unit energy that are available at each given position and on the probability $|\psi(\mathbf{r})|^2$ to find an electron at that given position. Thus, dI/dV is expected to show up peaks whenever the value of the bias equals the energy difference between the chain state and the tip. The bias, in turn, keeps the sample at a potential that can be either positive or negative with respect to the tip. In the former case the tunneling current involves unoccupied chain states, while in the latter the occupied states are involved.

A first data set is reproduced in Fig. 2.9. It displays the measurement of the differential conductance dI/dV detected by an STM on a set of chains of gold atoms assembled with the method illustrated above. A few useful remarks are in order. First, the measurements have been performed at positive bias, thus the tunneling current involves unoccupied electronic states. The curve refer to the different chains, and have been obtained after positioning the tip at their center. Interestingly, in the bias region where the direct measurement on the substrate has no peaks, the measurement on the chains shows instead well defined peaks. Their position moves towards lower bias while the chain length increases. A peak in dI/dV occurs whenever the bias equals the energy of a nanostructure state. In other words, the study of the peaks infers information on the energy levels in the chain. In particular, the first peak that is met while increasing the bias, corresponds to the lowest, or fundamental, state of the chain. Fig. 2.10 displays the comparison between experimental and theoretical data for the energy of the fundamental state as a function of the number of atoms in the chain. Squares represent the experimental data as they are extracted from Fig. 2.9. Circles represent the theoretical data as calculated from the Tight-Binding method discussed in Sec. 2.2.3, here expressed by the better suited relation

$$E = E_0 + 2\gamma \cos\left(\frac{\nu\pi}{N+1}\right). \quad (2.47)$$

Parameters E_0 and γ are fitted so to reproduce the experimental data. The measurement on a single atom in Fig. 2.10 fits the value $E_0 = 1.95$ eV, while the matrix element γ is evaluated from the data with $N > 1$. The value $\gamma = -0.62$ eV is found, which corresponds to the fundamental state with $\nu = 1$ and provides an excellent agreement with the experimental data.

An interesting finding is finally in order. The energy decreases with increasing the chain length. For $\gamma < 0$ and large values of N , (2.47) is approximated by

$$E_\nu = E_0 - 2|\gamma| + \frac{\pi^2 \nu^2}{(N+1)^2} |\gamma|. \quad (2.48)$$

This expression explicitly shows the decreasing behavior of the energy levels with increasing the length of the chain.

On physical grounds, consider an electron confined on a segment with length $L = Na$, so that its energy is completely of kinetic origin. The problem is similar

to that of a quantum particle in a 1D box with vanishing boundary conditions. It is known that the energy levels of such a particle are represented by

$$E_v = \frac{\pi^2 v^2}{L^2} \frac{\hbar^2}{2m} . \quad (2.49)$$

By inspection, (2.48) and (2.49) have identical structures: except the immaterial offset $E_0 - 2|\gamma|$, which represents the energy of the band minimum, the two expressions are identical if one assumes that the electron on the chain have the effective mass

$$m^* = \frac{\hbar^2}{2|\gamma|a^2} . \quad (2.50)$$

Inserting the values $a = 2.89 \text{ \AA}$ for the spacing between two gold atoms on the NiAl substrate, and $\gamma = -0.62 \text{ eV}$ as fitted from the experimental data with (2.47), one finds $m^* = 0.73m$: the effective mass turns out to be smaller than the rest mass of a free electron.

To conclude the illustration of the present application, one might argue that the Tight-Binding is capable of reproducing well enough the experimental data by means of only two parameters, in spite of the overall problem complexity. In fact, the matrix elements E_0 and γ embody the contributions coming from the electronic

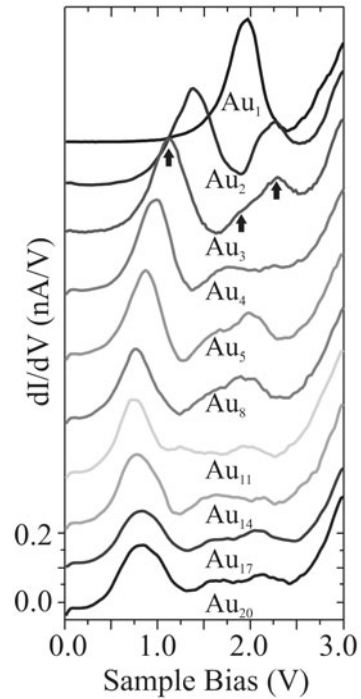


Fig. 2.9 Applications of the electronic structure concept to chains of metallic atoms. Differential conductance dI/dV obtained by STS as a function of applied bias, for selected gold chains on a NiAl substrate. For the sake of clarity, curves referring to different chains are disaggregated by rigidly offsetting each of them with respect to the others [7]

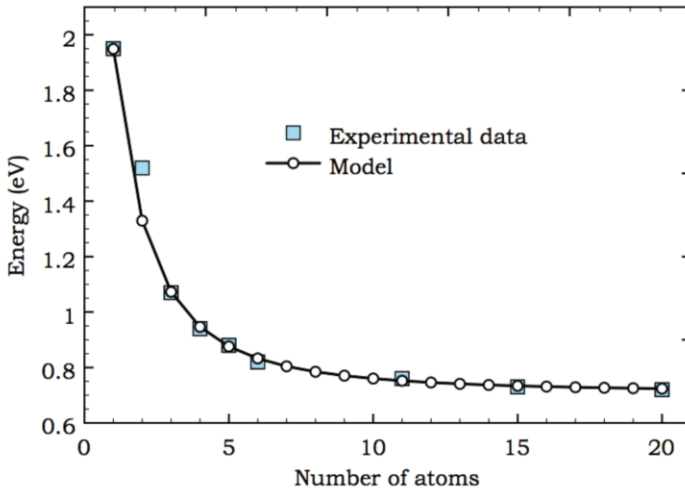


Fig. 2.10 Applications of electronic structure concept to chains of metallic atoms. Energy of the fundamental state for gold chains on a NiAl substrate, as a function of the number of atoms in the chain. Squares: experimental data as extracted from Fig. 2.9. Circles: theoretical data as calculated from fitting E_0 and γ to expression (2.47), derived by means of the Tight-Binding method

structure of both isolated gold atoms and of substrate atoms. Indeed, if E_0 and γ were to be calculated from first principles, one should necessarily include the substrate as well, trying to separate the two contributions.

2.2.8 Evolving from nanocrystals to infinite crystals

Sec. 2.2.3 has demonstrated that the energy levels of a one-dimensional nanocrystal with a large number of atoms, arrange themselves by thickening within a range of allowed energies, from there on referred to as energy bands. This issue is now going to be further explored.

On phenomenological grounds, consider for example a nanocrystal composed of silicon atoms as in Fig. 1.35. Besides the techniques exploited to synthesize the nanostructure, it is important to notice that the structure is quite similar, almost identical to that of a Si crystal. The image shows that the nanocrystal forms in an almost spherical geometry. Crystal planes are distinguishable in the inner part, evolve into an amorphous shape close to the surface, with the substrate on the outer part. The nanocrystal radius can be evaluated from the number of atoms and the lattice spacing $a = 0.357$ nm. Since the elementary cell of silicon contains two atoms and has the volume $a^3/4$, each atom occupies the volume $a^3/8$. In a sphere of radius R , the number N_{Si} of Si atoms is

$$\frac{4\pi}{3}R^3 = \frac{a^3}{8}N_{Si}.$$

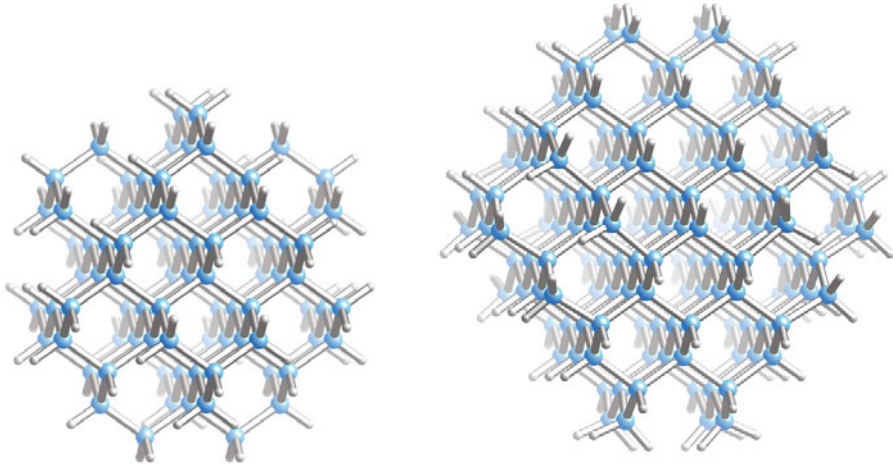


Fig. 2.11 Evolution from nano to infinite crystals. Left panel: a nanocrystal composed of 87 silicon atoms after cutting 76 bonds, that are then saturated by 76 hydrogen atoms. Right panel: sketch of the $\text{Si}_{147}\text{H}_{100}$ nanocrystal taken as example for the estimate in the text, showing the increasing complexity of electronic structure computations

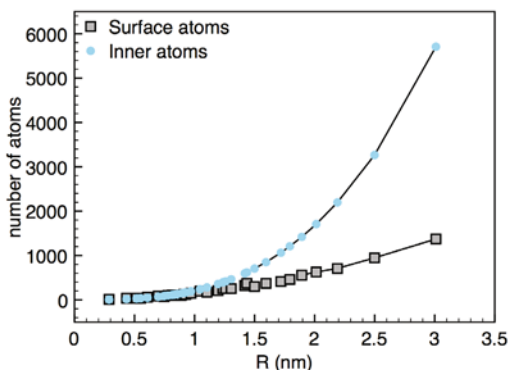
The value of R can be computed in nm as

$$R = \left(\frac{3a^3}{32\pi} N_{\text{Si}} \right)^{\frac{1}{3}} = 0.167 N_{\text{Si}}^{\frac{1}{3}} \text{ nm} , \quad (2.51)$$

The left panel of Fig. 2.11 displays a nanocrystal built up from a Si crystal with 87 atoms, after cutting 76 bonds with a radius of about 0.74 nm. The 76 pending bonds are saturated with an equal number of hydrogen atoms, represented by white small spheres in the figure. In real nanocrystals however, the surface can unlikely be terminated entirely by hydrogen atoms. In most cases, surfaces are composed of an oxide layer (here, SiO_2): this is either native, occurring since the nanocrystal formation and corresponding to the substrate, or it develops later on by exposure to air.

The relevance of surface states is depressed while the nanocrystal evolves towards the crystal. As seen in Sec. 2.2.5, matrix elements on surface sites might lead to formation of surface states, whose energy remains unchanged after variations of the structure size (see Figs. 2.7 and 2.8). These surface states may affect optical and transport properties. Internal states instead, those determining the energy bands while the nanostructure evolves towards the crystal, are not significantly affected by the surface states. Indeed, if the volume of the nanocrystal is increased after adding atoms and ordering them into the crystalline positions, the ratio of the number of internal to external atoms increases. Fig. 2.12 shows the number of surface and volume atoms as functions of sphere radius R . The surface plays a significant role for

Fig. 2.12 Evolution from nano to infinite crystals. Nanocrystal of Si with spherical geometry. The number of internal (circles) and surface (squares) atoms as functions of the sphere radius R



quite small nanocrystals, with radii amounting to no more than $1.0 \div 1.5$ nm. For larger nanocrystals the presence of the surface becomes progressively negligible.

It has been seen that while the nanostructure size is increased, electronic structures evolve from a sequence of single energy levels into an energy band. This is not the sole manifestation of the evolution from nano to infinite crystals. Optical and transport properties are as well affected. As an example, the coefficient of optical absorption in a nanocrystal is structured into peaks, whose height and energy positions depend on the nanocrystal size. The peak heights scale with the size, so that after a given threshold the overall lineshape of the absorption coefficient cannot be distinguished from that of a macroscopic crystal. One reason why nanostructures are so interesting for applications, resides precisely on the possibility of tailoring the optical properties after fine tuning of the size. This important issue is detailed in Chaps. 5 and 6.

A similar reasoning holds for transport properties. As an example, the phenomenological Ohm's law states that the electrical resistance of a macroscopic crystal scales with its length and it is inversely proportional to its transverse section, at least in a sample with cylindrical geometry. If the crystal length is shortened, the resistance decreases until it reaches a minimum value that is independent of the crystal nature. This is confirmed by experiments. Chap. 4 details this behavior and shows that these manifestations can be traced back to the modifications of the energy levels and of diffusion mechanisms after a change of crystal size.

As to the behavior of surface and impurity states with increasing the nanostructure size, Sec. 2.2.5 has shown that modifications of the parameters describing the end atoms in finite linear chains lead to the formation of surface states, whose energy and localization does not depend on the chain length. If two identical chains were connected to each other, the contact atom would be characterized by parameters identical to those of the internal atoms. However, if these parameters are different, an impurity state arises.

A general interesting issue is whether a critical size exists, below which the crystal acquires a "nano" behavior. There is no unique answer to this question, though intuition suggests that for each different physical property, a threshold for nano be-

havior would be dictated by a typical energy or length associated to that property. For electric conduction, the mean free path that an electron travels between two successive interactions is a possible typical length λ . In a macroscopic crystal, its order of magnitude can be as large as several tens of nm. If the nanocrystal size is smaller than λ , a nano behavior would presumably show up in electric conductivity. In any event, λ depends on temperature and purity degree, so that it is not possible to define a neat threshold between a nano and a macro behavior. Finally, for properties different than electrical conductivity, the length scale can be different.

To conclude, an estimate is now given of the increasing complexity in the computation of electronic states, while the nanocrystal size becomes larger. This point is tackled by the following example. Assume that the task were to compute by TB method the electronic energy levels of the nanocrystal $\text{Si}_{147}\text{H}_{100}$ in the right panel of Fig. 2.11, whose radius estimated from (2.51) be about 0.88 nm. To build the wavefunctions (2.1), one might limit the basis set to one s and three p states for each Si atom, and one s state for each hydrogen. The size of the matrix to be diagonalized rapidly becomes as large as $4 \times 147 + 100 = 688$ times 688. A computer and a good diagonalization algorithm may do the job. Nanocrystals larger than such a size can hardly be handled, since diagonalization algorithms are characterized by execution times that scale with the cube of matrix sizes. A nanocrystal of $\text{Si}_{705}\text{H}_{300}$, whose radius is about 2.5 nm, would require execution times about 100 times longer than in the previous case. Computation of electronic structures of macroscopic crystals by exact diagonalization cannot therefore be handled. The case of the one-dimensional atomic chain discussed in Sec. 2.2.3 is especially easy, since the matrix could be diagonalized by analytical methods whatever the size could have been. However, if matrix elements between second, third and so on neighbors had to be included, no analytical solution could be elaborated. These simple considerations lead to the conclusion that studies of large sized nanocrystals as well as macroscopic crystals need the introductions of new suited tools. One conceptual step forward is the subject of discussion in the next section.

2.3 Electronic structure of periodic crystals

A power conceptual tool exists, which opens the way to the solution of the problem and potentially overcomes the computational complexity remarked in Sec. 2.2.8. A crystal is certainly a system with macroscopic size and with a large though finite number of atoms. Consider the limit where the surface effects can be neglected, in the sense discussed in Sec. 2.2.8, and the crystal can be viewed as if it were infinitely extended along the three spatial directions. Under these conditions, the crystal would possess the property of being invariant under lattice translations. Thus, for each \mathbf{R}_n the following relations

$$H(\mathbf{r}) = H(\mathbf{r} + \mathbf{R}_n) \quad \forall \mathbf{R}_n \quad (2.52)$$

hold. Translational symmetry dictates suited restrictions on the form that the wavefunctions might assume. The general idea is that translational invariance could limit the computation of eigenstates to within the primitive cell. Charge density is for example a periodic function with periodicity dictated by the Bravais lattice, and the same holds for the potential energy. Periodicity of the charge density does not necessarily reflect on the wavefunctions though. Along these lines, the minimal true statement at hand remains that if $\psi(\mathbf{r})$ is eigenstate of $H(\mathbf{r})$, therefore $\psi(\mathbf{r} + \mathbf{R}_n)$ is eigenstate of $H(\mathbf{r} + \mathbf{R}_n)$ with one and the same energy, though the two ψ wavefunctions might differ by a phase from one another. The Bloch theorem is the powerful conceptual tool assisting in the solution of this problem, and provides the calculation of the phase along the following lines.

2.3.1 Bloch theorem

The lattice translation operators T_n are labeled by a site index n , defined by the relation

$$T_n \psi(\mathbf{r}) = \psi(\mathbf{r} + \mathbf{R}_n) . \quad (2.53)$$

They satisfy commutation relations $[T_n, T_{n'}] = 0$ among themselves, and with H . Indeed, one has:

$$T_n H(\mathbf{r}) \psi(\mathbf{r}) = H(\mathbf{r} + \mathbf{R}_n) \psi(\mathbf{r} + \mathbf{R}_n) = H(\mathbf{r}) \psi(\mathbf{r} + \mathbf{R}_n) = H(\mathbf{r}) T_n \psi(\mathbf{r}) .$$

These commutation properties imply that operators T_n and H share the same eigenfunctions, that is

$$\begin{aligned} H \psi(\mathbf{r}) &= E \psi(\mathbf{r}), \\ T_n \psi(\mathbf{r}) &= C(\mathbf{R}_n) \psi(\mathbf{r}). \end{aligned} \quad (2.54)$$

Now, the charge density is the squared modulus of the electronic wave function. Since the charge density is periodic, the squared modulus of the wavefunction has to remain unchanged after translations. That is:

$$|T_n \psi(\mathbf{r})|^2 = |\psi(\mathbf{r})|^2 . \quad (2.55)$$

One may thus conclude that

$$|C(\mathbf{R}_n)|^2 = 1 , \quad (2.56)$$

namely that eigenvalues of the translation operator must have a unit modulus square. This suggests that $C(\mathbf{R}_n)$ can be written in the form

$$C(\mathbf{R}_n) = e^{i\mathbf{k} \cdot \mathbf{R}_n} , \quad (2.57)$$

where \mathbf{k} is a real wavevector. The fundamental result is then derived:

$$T_n \psi(\mathbf{r}) = \psi(\mathbf{r} + \mathbf{R}_n) = e^{i\mathbf{k} \cdot \mathbf{R}_n} \psi(\mathbf{r}) . \quad (2.58)$$

Equation (2.58) expresses the Bloch theorem. It states that

Concept

The phase accumulated by the electronic wavefunction after a translation by \mathbf{R}_n is as simple as $\mathbf{k} \cdot \mathbf{R}_n$, with \mathbf{k} the wavevector and \mathbf{R}_n the lattice vector.

An alternative proof of the theorem can be given, bringing additional physical content. Equation (2.53) implies $T_n T_{n'} \psi(\mathbf{r}) = \psi(\mathbf{r} + \mathbf{R}_n + \mathbf{R}_{n'}) = T_{n+n'} \psi(\mathbf{r})$. Thus, $C(\mathbf{R}_n)$ has to be of the form (2.57), where this time \mathbf{k} is to be complex. If however \mathbf{k} had an imaginary part, the wavefunction $\psi(\mathbf{r})$ would be physically meaningless: the norm of $\psi(\mathbf{r})$ indeed, would not be finite nor scaling as the system volume. As a result, since $|\psi(\mathbf{r})|$ is related to the charge density, the charge density would be exponentially increasing or decreasing while passing from one cell or the other. This behavior conflicts with translational invariance.

Conversely,

Concept

Whenever translational invariance is corrupted for whatever reason, solutions with complex \mathbf{k} are admitted, which represent localized wavefunctions.

A similar situation has indeed been met while studying surface and impurity states in Sec. 2.2.5. Further examples of localized wave functions are investigated later on, when lattice defects are considered which break translational invariance as well.

In conclusion, the consequence of a perfect, infinite and periodic crystal on electronic wavefunctions is expressed by (2.58) with \mathbf{k} real, representing a precise prescription on the change that may occur on an electronic wave function after a lattice translation is operated. Since a crystal is essentially a set of juxtaposed primitive cells, (2.58) in practice tells that

Concept

The electronic wavefunction can be determined just inside the primitive cell, not necessarily everywhere.

The translation of this statement in mathematical language is as follows. Multiply both left and right hand sides of (2.58) by $e^{-i\mathbf{k} \cdot (\mathbf{r} + \mathbf{R}_n)}$. Therefore

$$e^{-i\mathbf{k} \cdot (\mathbf{r} + \mathbf{R}_n)} \psi(\mathbf{r} + \mathbf{R}_n) = e^{-i\mathbf{k} \cdot \mathbf{r}} \psi(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r}) ,$$

with

$$u_{\mathbf{k}}(\mathbf{r} + \mathbf{R}_n) = u_{\mathbf{k}}(\mathbf{r}) .$$

Thus $\psi(\mathbf{r})$ can be written in a more expressive form:

Definition

The expression for $\psi(\mathbf{r})$:

$$\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}) , \quad (2.59)$$

with $u_{\mathbf{k}}(\mathbf{r})$ a periodic function with the Bravais lattice periodicity, defines the Bloch wavefunction.

It states in essence that

Concept

The wavefunction of an electron in a perfect crystal is a plane wave $e^{i\mathbf{k}\cdot\mathbf{r}}$ whose amplitude is modulated by a function $u_{\mathbf{k}}(\mathbf{r})$ that is periodic on the Bravais lattice. Determinations of $\psi(\mathbf{r})$ can thus be reduced to a calculation of $u_{\mathbf{k}}(\mathbf{r})$ just in the primitive cell (not everywhere in the lattice).

The consequences of such a concept on the computational complexity are evident, and are even more enlightened by looking at the Schrödinger equation

$$-\frac{\hbar^2}{2m} \nabla^2 \psi + V \psi = E \psi , \quad (2.60)$$

where V is the potential felt by the electrons. Substitution of (2.59) into (2.60) yields the following equation for $u_{\mathbf{k}}(\mathbf{r})$:

$$\left(-\frac{\hbar^2}{2m} \nabla^2 - \frac{i\hbar^2}{m} \mathbf{k} \cdot \nabla + \frac{\hbar^2 k^2}{2m} \right) u_{\mathbf{k}} + V u_{\mathbf{k}} = E u_{\mathbf{k}} . \quad (2.61)$$

This is the equation to be solved just in the primitive cell. The computational complexity is thus independent of the number of cells composing the crystal. The particular solution to this differential equation requires a choice for the boundary conditions, which are to reflect the periodicity $u_{\mathbf{k}}(\mathbf{r})$. This issue is discussed in detail in dedicated textbooks, such as [4, 5, 8].

2.3.2 Periodic boundary conditions

Bloch theorem works as a fundamental step forward to comprehend the structure of electronic states in crystals. However, its present form remains of limited use as a tool for realistic crystals, as those found in nature or synthesized in factories, like

nanostructures for electronic devices. For realistic crystals, translational invariance is broken by the finite geometry, thus the Bloch theorem is not valid. A way out has to be devised, to keep together translational invariance and finite crystal geometry. Otherwise the Bloch theorem, though powerful and extraordinary conceptual tool it might be, remains of little usefulness.

In a finite crystal, Bloch functions could be used along with the boundary conditions as follows. Consider a given energy E : all Bloch states can be taken such that $E = E_n(\mathbf{k}) \forall n$, then an arbitrary linear combination of them can be composed, whose coefficients are finally determined in a way to satisfy the boundary conditions. In a macroscopic crystal, on physical grounds one would expect that the results be independent of the boundary conditions. This prediction follows the findings on the surface states in Sec. 2.2.5. Now, the range of the interaction between electrons and atoms is finite, as it has been discussed on phenomenological grounds in Chap. 1 and it is demonstrated in Chaps. 4–6. For each given confined system, internal and surface atoms can thus be distinguished according to the interaction range. Internal atoms are those for which the interaction range with electrons is internal to the system, and surface atoms are those whose distance from the surface is shorter than the interaction range. If the number of surface atoms N_s is negligible with respect to the total number of atoms N_t , that is the condition $N_s \ll N_t$ is verified, one can state that the number of surface states is negligible with respect to the internal states, and surface atoms can be to first approximation disregarded. Since the boundary conditions modify the behavior of surface atoms, stating that physical results are independent of the boundary conditions appears to be a reasonable conclusion.

Consider a crystal composed of N_1 primitive cells along the direction of primitive vector \mathbf{a}_1 , and of N_2 and N_3 cells along the remaining two directions. The crystals can be made infinite and periodic by simply adding replicas along the three directions. As an example, consider the atomic chain with N atoms studied in Sec. 2.2.1 and add replicas as displayed in Fig. 2.13. In other words, infinite replicas are added to left and right sides of the original N -atoms chain, so that translational invariance be recovered. This construction imposes conditions on the wavefunctions, since all the cells whose positions differ by N lattice spacings are to be supposed equivalent. This results into periodic boundary conditions, that is:

$$\psi(\mathbf{r}) = \psi(\mathbf{r} + N\mathbf{a}\hat{\mathbf{x}}). \quad (2.62)$$

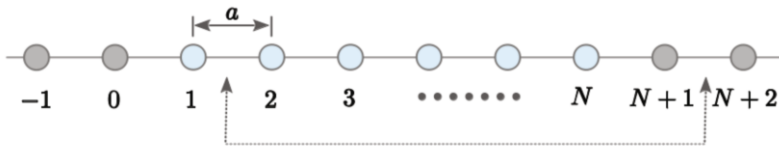


Fig. 2.13 Electronic structure of infinite crystals. An infinite periodic chain is originated by a finite crystal with N atoms (cyan spheres) periodically replicated infinite times (grey spheres). The two dotted arrows mark two cells that are equivalent due to periodic boundary conditions

On the other hand, the Bloch theorem (2.58) requires that

$$\psi(\mathbf{r} + Na\hat{\mathbf{x}}) = e^{ik \cdot (Na\hat{\mathbf{x}})} \psi(\mathbf{r})$$

and thus (2.62) must become

$$e^{ik \cdot (Na\hat{\mathbf{x}})} = 1 \rightarrow k = \frac{2\pi}{Na} v, \text{ with } v = 0, 1, \dots, N-1. \quad (2.63)$$

Expression (2.63) implies that periodic boundary conditions restrict the allowed values of \mathbf{k} .

Crystal periodicity and periodicity induced by periodic boundary conditions are close concepts, and mixing one with the other would lead to a conceptual mistake. Crystal periodicity is related to the primitive cell and implies the Bloch theorem, stating that not the whole $\psi(\mathbf{r})$ is periodic but only a part of it, $u_{\mathbf{k}}(\mathbf{r})$. Periodic boundary conditions are instead induced by the need of replicating a finite crystal over the whole space and require periodicity of $\psi(\mathbf{r})$ when stepping from one replica to another one. A subtle but important difference.

Extension to the case of a three dimensional crystals is simple. If N_1 is the number of primitive cells along \mathbf{a}_1 , and in analogous way N_2 and N_3 are introduced, (2.63) immediately implies:

$$e^{ik \cdot \mathbf{a}_i N_i} = 1, \text{ with } i=1, 2, 3.$$

Therefore

$$\mathbf{k} \cdot \mathbf{a}_i N_i = 2\pi v_i, \text{ with } i=1, 2, 3. \quad (2.64)$$

The set of vectors \mathbf{k} solving (2.64) are given by

$$\mathbf{k} = \frac{v_1}{N_1} \mathbf{b}_1 + \frac{v_2}{N_2} \mathbf{b}_2 + \frac{v_3}{N_3} \mathbf{b}_3, \quad (2.65)$$

where \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_3 are primitive vectors of the reciprocal lattice. In the previous one-dimensional examples of N -atoms chains, states have been labeled by an integer number v assuming N distinct values. The same holds here, where a finite system is considered, though with different boundary conditions. The integers v_i can be expected to assume the values $0, 1, 2, \dots, N_i - 1$. The total number of allowed \mathbf{k} vectors is $N_1 N_2 N_3$, that is the number of primitive cells in the whole crystal.

Consider now the procedure to apply the Bloch theorem and determine the crystal electronic structure. In one-dimensional cases, the problem is reduced to a diagonalization of the Hamiltonian matrix with size equal to the number N of atoms times the number of orbitals per atom. In the case of only one orbital per atom, the eigenvalue problem is

$$\sum_m H_{n,m} \phi_m = E \phi_n. \quad (2.66)$$

The coefficients ϕ_n determine $\psi(\mathbf{r})$ by means of (2.1). Now, instead of solving (2.66), Bloch theorem can be exploited to remind that

$$\phi_n = \phi_0 e^{ikna}, \quad (2.67)$$

with ϕ_0 referred to the site arbitrarily chosen as origin. Substitution of (2.67) into (2.66) yields:

$$\sum_m H_{n,m} \phi_0 e^{ikma} = E \phi_0 e^{ikna},$$

and thus

$$E(k) = \sum_m H_{n,m} e^{ik(m-n)a}. \quad (2.68)$$

In practice, Bloch theorem directly provides the eigenvalues, in the number of one for each allowed k value. One could remark that in fact the sum in (2.68) depends on site n . This would indeed be the case if the chain were finite. But if the chain is infinite and obtained by periodic replicas, one has $H_{n+p,m+p} = H_{n,m}$ for all integers p , thus $H_{n,m} = H_{0,m-n}$ and the sum becomes independent of n . Equation (2.68) thus turns into $E(k) = \sum_m H_{0,m} e^{ikma}$. Since $E(k)$ is a real function, one has $E(k) = \sum_m H_{0,m} [\cos(kma) + i \sin(kma)] = \sum_m H_{0,m} \cos(kma) + i \sum_m H_{0,m} \sin(kma)$ and thus $H_{0,m} = H_{0,-m}$. Indeed,

$$\begin{aligned} 0 &= \sum_m H_{0,m} \sin(kma) = \sum_{m \geq 1} H_{0,m} \sin(kma) + \sum_{m \leq -1} H_{0,m} \sin(kma) \\ &= \sum_{m \geq 1} (H_{0,m} - H_{0,-m}) \sin(kma). \end{aligned}$$

The energy levels dispersed in k -space to form energy bands, can actually be expressed as

$$E(k) = \sum_m H_{0,m} \cos(kma), \quad (2.69)$$

that is even in k . The following properties are seminal and quite useful:

Properties

P1. $E(k)$ is unchanged after translations of k -vector by integer multiples of $2\pi/a$, that is $E(k)$ is a periodic function of k .

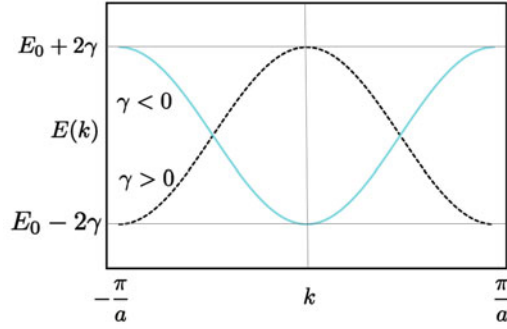
P2. $E(k)$ is unchanged after inversion of k -vector, $E(-k) = E(k)$, that is $E(k)$ is an even function of k .

P3. From (2.65) it descends that the total number of electronic levels in an energy band is two times the number of primitive cells in the crystal, the factor of two accounting for the spin degeneracy.

The former property P1 can be also expressed as follows. Since $\forall n \ G = 2\pi n/a$ are related to the reciprocal lattice vectors in the chains, one finds that

$$E(k + G) = E(k). \quad (2.70)$$

Fig. 2.14 Electronic structure of infinite crystals. Energy bands of a periodic chain represented within the first Brillouin zone. Solid line: case with $\gamma < 0$. Dashed line: case with $\gamma > 0$



The same property holds of course for the coefficients ϕ_n . It follows that two solutions differing by a reciprocal lattice vector are one and the same solution. This suggests that a minimal representation is obtained limiting the k vectors to the interval $-\pi/a < k \leq \pi/a$, that is the first Brillouin zone. The energy bands $E(k) = E_0 + 2\gamma\cos(ka)$, the energy bands, can thus be represented as Fig 2.14, the two representations being different according to the sign of γ .

Case with two atoms per cell. Consider now the more complex case of a chain with two atoms per cell, as in Fig. 2.15, where A and B label the atoms in the primitive cell and γ_1 and γ_2 are the matrix elements between first neighbors. For this case the Bloch theorem is expressed by explicitly writing (2.66) as

$$\begin{pmatrix} 1A & 1B & 2A & 2B & 3A & 3B & \cdots \\ 1A & \begin{pmatrix} E_0 & \gamma_1 \\ \gamma_1 & E_0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ \gamma_2 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & \cdots \\ 2A & \begin{pmatrix} 0 & \gamma_2 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} E_0 & \gamma_1 \\ \gamma_1 & E_0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ \gamma_2 & 0 \end{pmatrix} & \cdots \\ 3A & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & \gamma_2 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} E_0 & \gamma_1 \\ \gamma_1 & E_0 \end{pmatrix} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \begin{pmatrix} \phi_{1A} \\ \phi_{1B} \end{pmatrix} \\ \begin{pmatrix} \phi_{2A} \\ \phi_{2B} \end{pmatrix} \\ \begin{pmatrix} \phi_{3A} \\ \phi_{3B} \end{pmatrix} \\ \vdots \end{pmatrix} = E \begin{pmatrix} \begin{pmatrix} \phi_{1A} \\ \phi_{1B} \end{pmatrix} \\ \begin{pmatrix} \phi_{2A} \\ \phi_{2B} \end{pmatrix} \\ \begin{pmatrix} \phi_{3A} \\ \phi_{3B} \end{pmatrix} \\ \vdots \end{pmatrix} \quad (2.71)$$

The matrix structure helps understanding how the Bloch theorem is best used here. To orient the comprehension, consider for one moment the problem of only one atom per cell, when the matrix structure is:

$$\begin{pmatrix} E_0 & \gamma & 0 & \cdots \cdots \\ \gamma & E_0 & \gamma & 0 & \cdots \\ 0 & \gamma & E_0 & \gamma & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \end{pmatrix} = E \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \end{pmatrix} \quad (2.72)$$

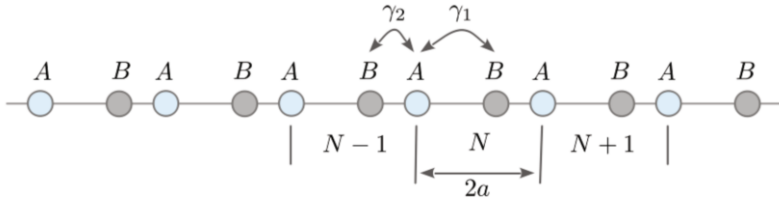


Fig. 2.15 Electronic structure of infinite crystals. Chain with two atoms per cell. Each primitive cell has two atoms, labeled A and B. Also indicated are the matrix elements between first neighbors, γ_1 and γ_2

Comparison between (2.71) and (2.72) makes clear that the introduction of a second atom per cell transforms single matrix elements in (2.72) into 2×2 blocks as in (2.71). Expansion of (2.71) indeed yields:

$$H_{n,n} \{\phi\}_n + H_{n,n-1} \{\phi\}_{n-1} + H_{n,n+1} \{\phi\}_{n+1} = E \{\phi\}_n, \quad (2.73)$$

where

$$H_{n,n} = \begin{pmatrix} E_0 & \gamma_1 \\ \gamma_1 & E_0 \end{pmatrix}, H_{n,n-1} = \begin{pmatrix} 0 & \gamma_2 \\ 0 & 0 \end{pmatrix}, H_{n,n+1} = \begin{pmatrix} 0 & 0 \\ \gamma_2 & 0 \end{pmatrix}, \{\phi\}_n = \begin{pmatrix} \phi_A \\ \phi_B \end{pmatrix}_n. \quad (2.74)$$

On formal grounds, (2.73) is similar to (2.8), though the repeated structural unit has become a pair, where before it was related to a single atom. In analogy with (2.67), the Bloch theorem therefore can be expressed as:

$$\begin{pmatrix} \phi_A \\ \phi_B \end{pmatrix}_n = \begin{pmatrix} \phi_A \\ \phi_B \end{pmatrix} e^{2ikna}, \quad (2.75)$$

by considering, as in Fig. 2.15, a lattice spacing of $2a$. Substitution of (2.75) into (2.73) leads to

$$\left[\begin{pmatrix} E_0 & \gamma_1 \\ \gamma_1 & E_0 \end{pmatrix} + \begin{pmatrix} 0 & \gamma_2 \\ 0 & 0 \end{pmatrix} e^{-2ika} + \begin{pmatrix} 0 & 0 \\ \gamma_2 & 0 \end{pmatrix} e^{2ika} \right] \begin{pmatrix} \phi_A \\ \phi_B \end{pmatrix} = E \begin{pmatrix} \phi_A \\ \phi_B \end{pmatrix}. \quad (2.76)$$

Summing the three matrices on the left-hand side of (2.76), the simplified matrix equation is obtained:

$$\begin{pmatrix} E_0 & \gamma_1 + \gamma_2 e^{-2ika} \\ \gamma_1 + \gamma_2 e^{2ika} & E_0 \end{pmatrix} \begin{pmatrix} \phi_A \\ \phi_B \end{pmatrix} = E \begin{pmatrix} \phi_A \\ \phi_B \end{pmatrix}. \quad (2.77)$$

Thus, determination of the electronic structure of a chain with two atoms per cell requires the diagonalization of a 2×2 matrix instead of (2.71).

Generalization of the above procedure to a generic periodic crystal is conceptually obvious:

Procedure

Step 1. Determine the size of the original matrix from the number of primitive cells and of the base atomic orbitals in the cell. If N is the number of primitive cells and N_b the number of orbitals of the base in the cell, the starting matrix has the gigantic size $(NN_b) \times (NN_b)$.

Step 2. Apply the approximation of periodic boundary conditions, transforming the finite into an infinite crystal.

Step 3. Write H and ϕ in blocks. As a consequence of Step 2, these turn out to be as small as $N_b \times N_b$ and $N_b \times 1$, respectively.

Step 4. The eigenvalues problem is thus

$$\sum_m [H]_{n,m} \{\phi\}_m = E \{\phi\}_n . \quad (2.78)$$

As in (2.73), index m and n refer to the cell and each of $[H]_{n,m}$ matrices has the size $N_b \times N_b$.

Step 5. Use Bloch theorem and $H_{n,m} = H_{0,m-n}$ to transform expression (2.78) into

$$\sum_m [H]_{0,m-n} \{\phi\}_0 e^{i\mathbf{k} \cdot \mathbf{R}_m} = E \{\phi\}_0 e^{i\mathbf{k} \cdot \mathbf{R}_n} .$$

Step 6. Define the matrix $H(\mathbf{k})$ by means of

$$H(\mathbf{k}) = \sum_m [H]_{0,m-n} e^{i\mathbf{k} \cdot (\mathbf{R}_m - \mathbf{R}_n)} . \quad (2.79)$$

Step 7. The original eigenvalue problem (2.78) is then transformed into the simpler

$$H(\mathbf{k}) \{\phi\}_0 = E \{\phi\}_0 . \quad (2.80)$$

Step 8. Solve the eigenvalue problem (2.80) with respect to eigenenergies E and eigenstates ϕ .

The first part of the above procedure is now applied to the following example.

Examples

Suppose to determine the band structure of silicon starting from a base of four orbitals per cell, one with s and three with p symmetry. Silicon has two atoms per

primitive cell. Four orbitals for each of the two atoms amount to $N_b = 8$ elements of the base. Thus the starting matrix is $8N \times 8N$ sized, with N being the number of primitive cells (of the order of Avogadro number, to have an idea). Use of Bloch theorem reduces the problem to the diagonalization of an 8×8 matrix, for each \mathbf{k} inside the first Brillouin zone. Since the number of \mathbf{k} vectors allowed by periodic boundary conditions is N , that is the number of primitive cells, the number of eigenvalues eventually found after diagonalization amounts to $8N$. From a computational point of view, it is enormously simpler to diagonalize N times an 8×8 matrix rather than one time an $8N \times 8N$ matrix. Of course, this is the case if the matrix is in blocks as shown above, and cross-blocks terms connecting states with different \mathbf{k} -vectors are negligible.

Remark on periodic boundary conditions. A final remark on periodic boundary conditions is in order. As already stated, the infinite crystal has been usefully obtained via infinite replicas of a finite crystal, so to fill the whole space. Considering a finite crystal composed of $N_1 N_2 N_3$ primitive cells and make it infinite is an useful procedure. The validity of this approximation is to be well defined and reminded. First, a finite crystal could realistically have surfaces contaminated by impurities composed of atoms of species different than the crystal one. The atoms on the surface sites would likely occupy positions different than those that they occupy inside the crystal. Besides, surface states might occur, as discussed in Sec. 2.2.5. Thus, it must be clarified that whenever periodic boundary conditions are used to transform a finite into an infinite crystal, one is assuming that the finite crystal is perfect, namely free from defects or impurities inside and on its surface as well, and that the number of cells be $N \gg 1$. Second, the approximation of infinite crystal is well posed and provides reliable results whenever the interest is on bulk crystal properties. Whenever surface or interface properties are instead concerned, periodic boundary conditions are no longer valid, at least in direction perpendicular to the surface or interface. On the other hand, peculiar and significant cases of finite crystals exist, where periodic boundary conditions are precisely valid, not even an approximation. Carbon nanotubes are one such example. Though the real growth process is a bit elaborated, carbon nanotubes can be thought as a single atomic layer of graphene that is cut and wrapped up, so to make the two edges coinciding. The wrapping up process is schematically shown in Fig. 2.16.

2.3.3 *Forbidden and allowed bands*

The example of a periodic chain with two atoms per cell is pedagogically useful to enlighten a general property of bands. Be $2a$ the lattice spacing of the periodic chain with two atoms per cell. The distance between the two atoms can be any value d in the interval $0 < d < a$. The eigenvalues calculated from (2.77) are

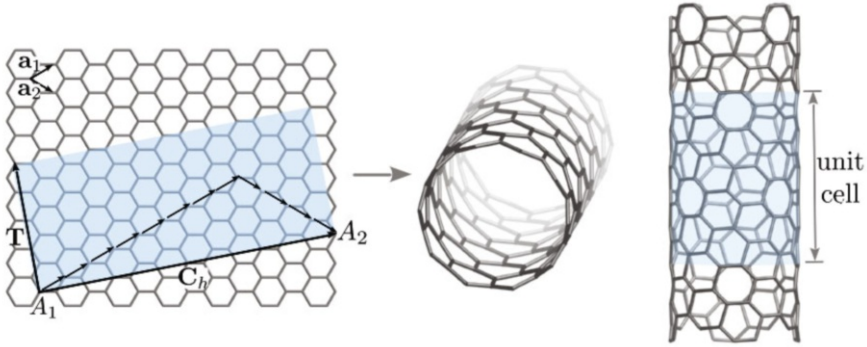


Fig. 2.16 Electronic structure of infinite crystals. Schematic wrapping up process leading to the formation of a carbon nanotube, evidencing a case of a finite crystal where periodic boundary conditions are not an approximation, as detailed in Sec. 2.3.4. The case of an (8,4) carbon nanotube is shown as an example. Rightmost panel: the cyan-shaded region highlights the carbon nanotube unit cell

$$E(k) = E_0 \mp \sqrt{(\gamma_1 - \gamma_2)^2 + 4\gamma_1 \gamma_2 \cos^2(ka)}. \quad (2.81)$$

A few interesting remarks are in order. First, the function $\cos^2(ka + p\pi) = \cos^2(ka)$ is periodic with period π , thus so are the functions $E(k)$ in (2.81), identifying the Brillouin zone with the interval $[-\pi/(2a), \pi/(2a)]$. That is, the functions $E(k)$ are periodic and even in k , here with period π/a , as previously observed.

Second, if $\gamma_1 = \gamma_2$, the result of (2.11) for a monoatomic chain is recovered. Setting $\gamma_1 = \gamma_2 = \gamma$ indeed, (2.81) becomes

$$\begin{aligned} E_1(k) &= E_0 - 2|\gamma| \cos(ka) \\ E_2(k) &= E_0 + 2|\gamma| \cos(ka). \end{aligned} \quad (2.82)$$

Fig. 2.17 displays the two bands $E_1(k)$ and $E_2(k)$ in the first Brillouin zone. The results (2.11) and (2.82) are shown to be equivalent after folding the expression (2.11) into the first Brillouin zone of the chain with two atoms per cell. Here is how the procedure works.

Procedure

Consider $E(k)$ given by Eq. (2.11) for each $k \in [-\pi/a, \pi/a]$. This is represented by the partly solid and partly dashed line in Fig. 2.17. The dashed part of the line is then folded back into the Brillouin zone $[-\pi/(2a), \pi/(2a)]$ as shown in figure, that is after a rigid shift by π/a towards either left (for the dashed line on the right) or right (for the dashed line on the left). The resulting band structure, now represented by two solid lines, corresponds to the two branches in (2.82).

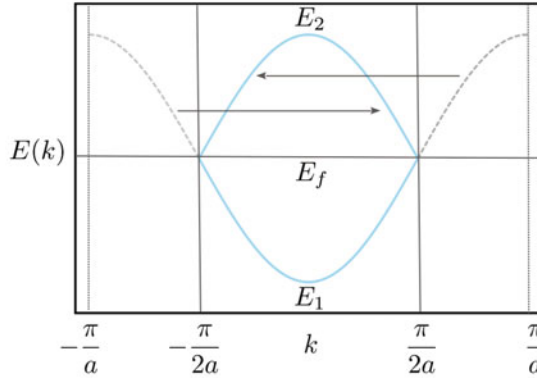


Fig. 2.17 Forbidden and allowed bands. The folding procedure aimed to represent the whole band structure into the first Brillouin zone is sketched (see text). Notice that the first Brillouin zone of the monoatomic chain is twice the first Brillouin zone of the biatomic cell

Third, if the coupling constants are different, corresponding to the situation depicted in Fig. 2.15, the band structure undergoes a qualitative change. The most important change occurs at the Brillouin zone edges, that is at $k = \pm\pi/(2a)$, where an energy gap opens up. Such a circumstance is displayed in Fig. 2.18 along with the energy values that are obtained from (2.81) at $k = 0$ and $k = \pm\pi/(2a)$.

Concept

A gap energy $E_g = 2|\gamma_1 - \gamma_2|$ opens up at the Brillouin zone boundaries as a result of the difference $|\gamma_1 - \gamma_2|$. Within the gap no solutions to the Bloch problem exist. Thus, that part of the band is forbidden to electrons.

Fourth, the band structure in Fig. 2.18 shows selected k -values in correspondence of which the first derivative of $E_1(k)$ and $E_2(k)$ with respect to k vanishes. In the figure, this occurs at $k = 0$ and $k = \pm\pi/(2a)$. Points with this property are generally referred to as critical points. An expansion of $E(k)$ in power series with respect to k and in the vicinity of these critical points, would show the term proportional to k^2 as the first non-zero term. This finding suggests to define the effective mass from the coefficient of the k^2 term in the expansion, that is from $E(k) = C + \hbar^2 k^2 / (2m^*)$. Considering first the band

$$E_1(k) = E_0 - \sqrt{(\gamma_1 - \gamma_2)^2 + 4\gamma_1\gamma_2 \cos^2(ka)},$$

and approximating $\sqrt{a + b \cos^2(x)} \simeq \sqrt{a + b} - bx^2 / (2\sqrt{a + b})$ around $x = 0$, one finds:

$$E_1(k) = E_0 - |\gamma_1 + \gamma_2| + \frac{2|\gamma_1\gamma_2|a^2}{|\gamma_1 + \gamma_2|} k^2. \quad (2.83)$$

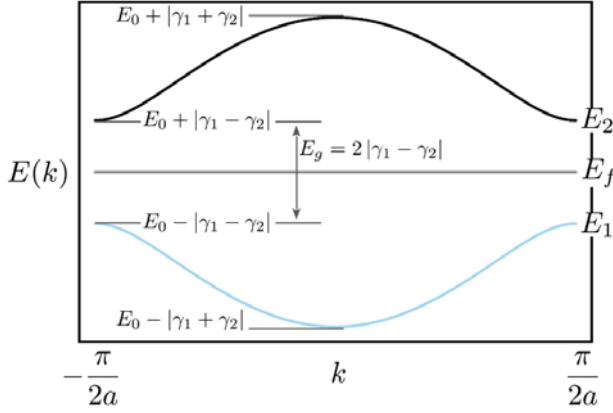


Fig. 2.18 Forbidden and allowed bands. The gap opening at the Brillouin zone boundaries is depicted. The size of the gap is indicated (see text)

The resulting effective mass turns out to be

$$m_0^* = \frac{\hbar^2}{4|\gamma_1 \gamma_2|a^2} |\gamma_1 + \gamma_2|. \quad (2.84)$$

With a similar procedure, approximating this time

$$\sqrt{a + b \cos^2(x)} \simeq \sqrt{a} + b/(2\sqrt{a}) (x - \pi/2)^2$$

around $x = \pi/2$, one obtains

$$E_1(k) = E_0 - |\gamma_1 - \gamma_2| - \frac{2|\gamma_1 \gamma_2|a^2}{|\gamma_1 - \gamma_2|} \left(k - \frac{\pi}{2a}\right)^2. \quad (2.85)$$

The effective mass results to be

$$m_{\pi/(2a)}^* = -\frac{\hbar^2}{4|\gamma_1 \gamma_2|a^2} |\gamma_1 - \gamma_2|. \quad (2.86)$$

Notice that in the case $\gamma_1 = \gamma_2 = \gamma$, the point $k = \pi/a$ ceases to be critical and $m_{\pi/(2a)}^*$ becomes meaningless. Actually, in this limit $m_{\pi/(2a)}^* = 0$. In both cases the effective mass depends on the gap at $k = 0$ and $k = \pi/(2a)$. Fig. 2.18 shows that at $k = 0$ the gap value is $E_g(0) = 2|\gamma_1 + \gamma_2|$, while at $k = \pi/(2a)$ the gap amounts to $E_g(\pi/(2a)) = 2|\gamma_1 - \gamma_2|$. In the vicinity of the critical point, the energy bands can be represented as

$$E_1(k) = E_0 - \frac{E_g(0)}{2} + \frac{\hbar^2}{2m_0^*} k^2 \quad \text{near } k = 0$$

$$E_1(k) = E_0 - \frac{E_g(\frac{\pi}{2a})}{2} + \frac{\hbar^2}{2m_{\pi/(2a)}^*} \left(k - \frac{\pi}{2a}\right)^2 \quad \text{near } k = \frac{\pi}{2a}$$

for $E_1(k)$, whereas for $E_2(k)$:

$$E_2(k) = E_0 + \frac{E_g(0)}{2} + \frac{\hbar^2}{2(-m_0^*)} k^2 \quad \text{near } k = 0$$

$$E_2(k) = E_0 + \frac{E_g(\frac{\pi}{2a})}{2} + \frac{\hbar^2}{2(-m_{\pi/(2a)}^*)} \left(k - \frac{\pi}{2a}\right)^2 \quad \text{near } k = \frac{\pi}{2a}.$$

From the above analysis it results that

Concept

The effective masses at a given critical point are in general different for each band and may have different signs.

For example, consider the case in which the wavefunction were expanded on a basis set which includes orbitals with p and possibly d symmetry, or else including neighbors farther than the first. In this case the effective masses of highest and lowest bands are different. These results are quite useful in applications.

Notice that the finite chain with two atoms per cell studied in Sec. 2.2.6 differs from the infinite chain with two atoms per cell studied in this section. In the former, the two atoms are distinguished via the on-site energy, which corresponds to atoms of different species but with uniform spacing. The resulting band structure is represented by (2.44). In the latter, instead, different values of first-neighbor interaction have been used, which corresponds to two different atomic species with non-uniform spacing, as sketched in Fig. 2.15. The corresponding band structure is expressed by (2.81). Problem 2.17 asks to elaborate on similarities and differences between the two problems.

In conclusion, the existence of an electronic band structure with its allowed and forbidden bands is in essence the consequence of the quantum nature of electrons: to the particle can indeed be associated a wave with well given boundary conditions that make it stationary. As all stationary waves, also electron standing waves in crystals have a fundamental mode and its overtones: these are the electronic energy bands. Given the electronic band structure with its allowed and forbidden bands, the question arises how these are filled in with the available electrons. While referring to Chapter 4 on transport properties for the detailed treatment, it is useful to state here the main idea that distinguishes metals from insulators and semiconductors. In a metal, the Fermi energy lies in the middle of an allowed band: therefore, carriers have available empty states where they can shift into under the action of an appropriate external field coupling to either their mass or charge, or spin. In an insulator instead, the Fermi energy lies in the middle of a band gap, so that carriers have no available empty states where to shift into, unless they are provided with a considerable energy to overcome the (large) gap and move them into an higher allowed

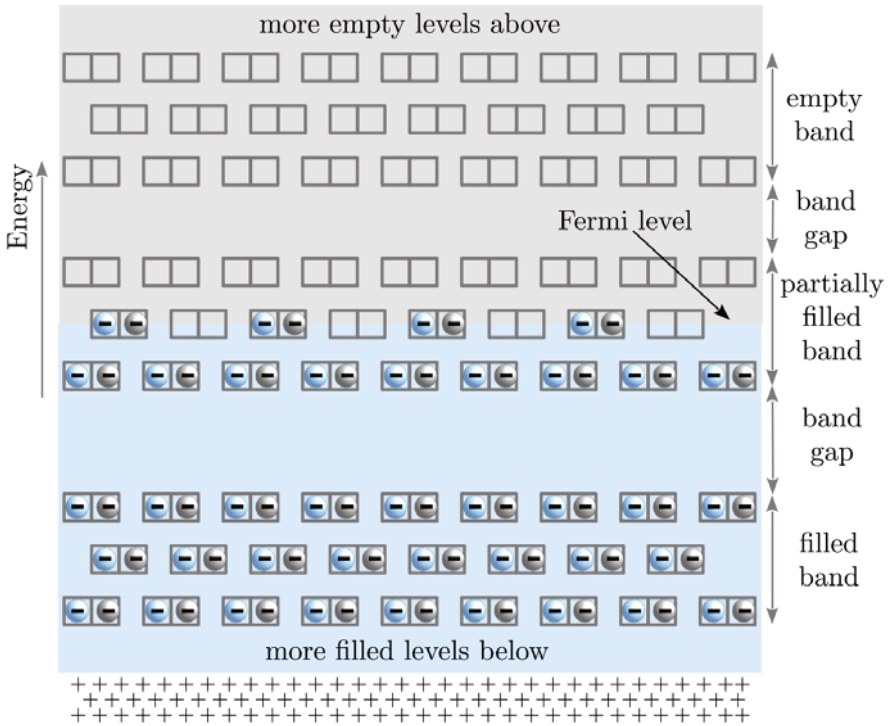


Fig. 2.19 Sketch of electronic levels in solids [1]. These are grouped in bands which are filled by electrons up to the Fermi level, according to the Pauli exclusion principle. Metals, semiconductors, and insulators are distinguished according to whether the Fermi energy lies in the middle of an allowed band as in a metal, or in the middle of a forbidden gap as in the case of a semiconductor or insulator. At the bottom, the positive charges represent the charged nuclei, ensuring charge neutrality

energy band. This can for example occur via photons when an insulator is stroke by suited incident light. In semiconductors, doping via addition of extra electrons or subtraction of present electrons in a (typically small gap) insulator, artificially frees empty states where the carriers can move into. Therefore, the semiconducting, insulating or metallic nature of a solids is intimately related with the filling of the bands, as schematically shown in Fig. 2.19.

Quick Questions

- Q5.** Refer to Fig. 2.17. If each atom has only one valence electron, where is the maximum energy of occupied states at zero temperature placed, that is the Fermi energy ϵ_f ?

Answer. Each electron has two possible spin states, and because of the Pauli principle, the Fermi energy ϵ_f is located midway within the band, so that there are unoccupied states with energy immediately above ϵ_f . The system turns out to have metallic properties, since electrons have unoccupied energy states where they can move into with no or tiny energy cost.

Q6. Refer to Fig. 2.18. If each atom has only one valence electron and considering Q5, tell whether the system can be considered to have metallic properties.

Answer. No. Electrons in this case occupy all the states of the lowest band. The first unoccupied state has energy E_g larger than the last occupied one. Thus, electrons have to spend a considerable amount of energy to move into unoccupied states.

Q7. Consider a one-dimensional system and suppose to have boundary conditions different from (2.62). In particular, $\psi(a) = \psi(Na) = 0$. Determine the allowed k -values.

Answer. Bloch theorem leads to

$$\begin{aligned} c_1 e^{ika} u_k(a) + c_2 e^{-ika} u_{-k}(a) &= 0 \\ c_1 e^{ikNa} u_k(Na) + c_2 e^{-ikNa} u_{-k}(Na) &= 0. \end{aligned}$$

Using $u_k(a) = u_k(Na)$ and $u_{-k}(a) = u_k^(a)$ one finds*

$$\begin{aligned} c_1 e^{ika} + c_2 e^{-ika} e^{i\lambda} &= 0 \\ c_1 e^{ikNa} + c_2 e^{-ikNa} e^{i\lambda} &= 0, \end{aligned}$$

with $e^{i\lambda} = u_k^(a)/u_k(a)$. The determinant of the homogeneous linear set is $e^{-ik(N-1)a} - e^{ik(N-1)a} = 2i \sin[k(N-1)a] = 0$, or else $k = p\pi/[(N-1)a]$, with p integer. Note that this has the same form as in (2.63).*

Q8. Consider a free particle with mass m^* and negative charge q in a constant electric field \mathbf{E} . Which direction does the particle velocity point at time t with respect to the direction of \mathbf{E} , if the particle is at rest at time $t = 0$? [Consider the particle as if it were moving according to classical equations of motion]

Answer. The particle velocity is along the direction of the electric field, according to $\mathbf{v} = (qt/m^)\mathbf{E}$. Thus, \mathbf{v} has the same direction and points towards \mathbf{E} if the signs of q and m^* are the same and opposite to \mathbf{E} if the signs are opposite.*

Q9. In laboratory, confined systems of a few thousands or million atoms are currently realized, which are at temperatures of a few tens of nK.

Under the lab conditions, these atoms behave as if they were either fermionic or bosonic particles subjected to periodic potentials provided by stationary waves of counterpropagating laser light. Would the Bloch theorem be valid for these systems?

Answer. Yes, it is, with atoms in place of electrons. Atoms surf the stationary light waves as a consequence of light-matter interactions. Actually such systems are to be considered analogous to perfect crystals, since no defects are by construction present. Considering the precisions that can be realized in atomic physics experiments, in principle there is not even an analogue of atoms vibrating around lattice sites.

Q10. The core of a xerographic copier machine is an inner belt made of a photoconductor material coated by a metallic layer [1]. A photoconductor is insulating under normal conditions but becomes a conductor when exposed to light, whose photons have sufficiently large frequency and thus energy. Schematically, light diffused by the copying document impinges on the photoconducting belt previously sprinkled with a lot of negative charges via a corona discharge mechanism, which by induction charges positively the metallic coating. The diffused light forms a pattern that is bright everywhere but in correspondence of text and images: the belt thus conducts only in the bright regions, freeing electrons that neutralize the positive charges in the metallic coating. When sprinkled, on-purpose positively charged toner particles are attracted by the remaining negative ones on the belt, forming an image of the original document. A second light exposure moves away the negative charges on the belt, leaving the positive toner charges loosely clinging. At this stage, the blank paper is passed over the belt after having being sprayed with negative charges, so that the toner particles are attracted onto the paper, where they are subsequently fused by heating and pressing processes. After charge erasing by light and cleaning, the cycle can start back [1]. Suppose that now you cover the original document with the transparent sheet that filters only the red component of light, stopping all the others. Would the copying process produce a reasonable result?

Answer. Probably not: red photons have the lowest-energy within the visible range, and presumably unable to make the photoconductor working, at least efficiently. Try and see...

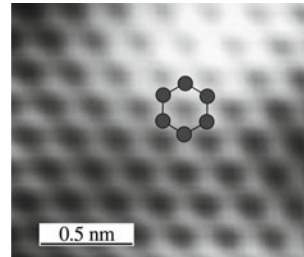
2.3.4 The effect of dimensionality through examples: the cases of graphene and carbon nanotubes

Electronic states of one-dimensional finite nanostructures and infinite crystals have been calculated so far, learning the basic concepts and tools on simple situations.

These are now applied to the calculation of the electronic structure of graphene (exfoliated graphite) and of carbon nanotubes. The choice of these two examples is aimed to enlighten that materials based on the same atomic species, carbon in this case, may have qualitatively relevant differences in their band structures and thus different properties.

Graphene. Graphene can be viewed either as a two-dimensional crystal or as a system with nanometric size. Fig. 2.20 displays an STM image of graphene, where the hexagonal structure is clearly visible from.

Fig. 2.20 Applications of electronic structure calculations to graphene. STM image of graphene, evidencing the hexagonal structure. Reprinted with permission from [9]. Copyright (2008) by the American Physical Society



The structure has already been investigated in Chap. 1, learning everything about positions, lattice and base. Carbon atoms occupy the vertex of regular hexagons so that each atom is surrounded by three else. The hexagon side corresponds to bond distance, that is about 1.4 \AA . In the example of Fig. 2.20, the substrate underlying the exfoliated graphite layers is composed of silicon carbide.

Electronic bands are now being calculated by means of Tight-Binding Method with first-neighbor approximation. First, the Bravais lattice has to be identified, in order to know the composition of the primitive cell. For this step, reference is made to Chap. 1. Besides knowing the atoms in the primitive cell, one has to infer the number of orbitals to be considered. Carbon has six electrons, corresponding to $1s^2 2s^2 2p^2$ structure, four of them being considered as valence electrons. When a carbon atom interacts with its three first neighbors, orbitals s , p_x , and p_y hybridize to form three bonds whose charge densities are localized as in Fig. 2.21. Orbital p_z instead, whose lobes are orthogonal to the graphene layer, hybridize among themselves to form π -like structures.

Therefore, graphene wavefunctions can be constructed as linear combinations of atomic orbitals s, p_x, p_y and p_z centered on the two atoms of the primitive cell. Six bands are then expected, originated from the sp^2 hybridization, and two bands from the π hybridization. In fact, accurate analysis of the matrix elements demonstrate that the whole 8×8 matrix reduces to two independent blocks, a 6×6 sp^2 block related to orbitals s, p_x, p_y and a 2×2 p_z block. This block separation, where the matrix elements between the two groups of atomic orbitals are zero, is a direct consequence of orthogonality of p_z orbitals to the graphene layer. If graphene were not perfectly planar, the Hamiltonian matrix would get non-zero cross elements and loose its block structure.

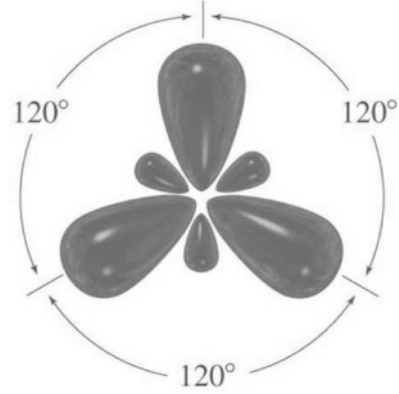


Fig. 2.21 Applications of electronic structure calculations to graphene. Schematic representation of sp^2 hybridization

For the sake of completeness, the elementary translation vectors, reciprocal vectors and first Brillouin zone are here reported again from Chap. 1, before proceeding to the calculation of the energy bands. The translation vectors are $\mathbf{a}_1 = a\hat{x} + b\hat{y}$ and $\mathbf{a}_2 = a\hat{x} - b\hat{y}$, with $a = 3a_0/2$ and $b = \sqrt{3}a_0/2$, with a_0 the first-neighbor distance. The reciprocal lattice vectors are: $\mathbf{b}_1 = (\pi/a)\hat{x} + (\pi/b)\hat{y}$ and $\mathbf{b}_2 = (\pi/a)\hat{x} - (\pi/b)\hat{y}$. The first Brillouin zone is sketched in the left panel of Fig. 2.22, along with the corresponding first-neighbor real-space cells structure in the right panel.

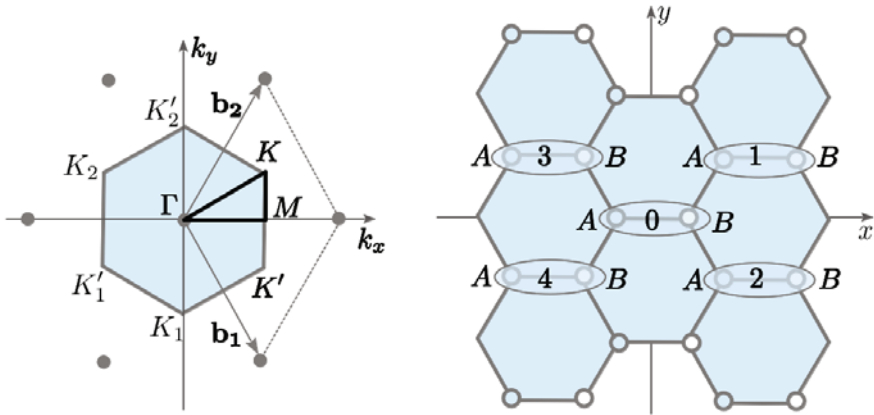
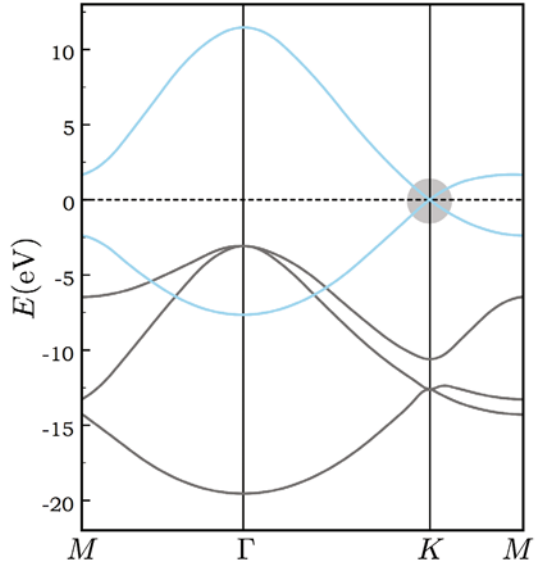


Fig. 2.22 Applications of electronic structure calculations. The case of graphene. Left panel: schematic representation of the first Brillouin zone represented by the shadowed region. The grey dots represent the sites of the reciprocal lattice generated by \mathbf{b}_1 and \mathbf{b}_2 (see text). The special symmetry points $\Gamma(0,0)$, $M(\pi/a,0)$ and $K(\pi/a, \pi/3b)$ are also reported. The K_i (K'_i) points with $i = 1$ and 2 , are equivalent to K (K') since they differ by a reciprocal lattice vector. Right panel: the four first neighboring cells entering the calculation of matrix $H(\mathbf{k})$ (see text), in real space

Fig. 2.23 Applications of electronic structure calculations. Band structure for the electronic states of graphene. The position of the Fermi level is represented by the dashed line, the grey circle indicating its intersection with the band structure



In Fig. 2.23 the complete band structure of graphene is displayed. In this spaghetti-like structure, a quite interesting part is the region around the Fermi level, represented by the dashed line. The grey circle evidences the point where the dashed line intersects the band structure. The two energy levels crossing the Fermi level correspond to the bands originated by p_z orbitals. In order to calculate them, the matrix $H(\mathbf{k})$ has to be first determined from (2.79). Since the problem has been reduced to two blocks, (2.76) can be effectively used with $\gamma_1 = \gamma_2 = \gamma$. Following the already illustrated procedure, one finds

$$[H]_{0,0} = \begin{matrix} 0A & 0B \\ 0A & 0B \end{matrix} \begin{pmatrix} E_0 & \gamma \\ \gamma & E_0 \end{pmatrix},$$

$$[H]_{0,1} = \begin{matrix} 1A & 1B \\ 0A & 0B \end{matrix} \begin{pmatrix} 0 & 0 \\ \gamma & 0 \end{pmatrix}.$$

A complication arises in this case because of the presence of terms related to the four structures in the right panel of Fig. 2.22. One has:

$$H(\mathbf{k}) = \begin{pmatrix} E_0 & \gamma f(\mathbf{k}) \\ \gamma f(\mathbf{k})^* & E_0 \end{pmatrix}, \quad (2.87)$$

with $f(\mathbf{k})$ the geometrical structure factor

$$\begin{aligned} f(\mathbf{k}) &= 1 + e^{-i\mathbf{k} \cdot \mathbf{a}_1} + e^{-i\mathbf{k} \cdot \mathbf{a}_2} \\ &= 1 + 2e^{ik_x a} \cos(k_y b). \end{aligned} \quad (2.88)$$

It is immediately seen that the eigenvalues of $H(\mathbf{k})$ for each \mathbf{k} in the first Brillouin zone are given by

$$E(\mathbf{k}) = E_0 \pm \gamma |f(\mathbf{k})| . \quad (2.89)$$

If needed, the fact that atomic orbitals centered on different sites are not orthogonal can be taken into account. In this case, besides $H(\mathbf{k})$ the matrix $S(\mathbf{k})$ has to be constructed, containing the overlaps. This can be done following the same procedure as for $H(\mathbf{k})$. After a few elaborations and using the same approximations as in Sec. 2.2.4, one finds

$$S(\mathbf{k}) = \begin{pmatrix} 1 & \sigma f(\mathbf{k}) \\ \sigma f^*(\mathbf{k}) & 1 \end{pmatrix} , \quad (2.90)$$

where σ is the overlap integral between first-neighbor sites. The eigenvalues calculated from (2.89) now become:

$$E(\mathbf{k}) = E_0 \pm \frac{\gamma |f(\mathbf{k})|}{1 \pm \sigma |f(\mathbf{k})|} . \quad (2.91)$$

The two bands obtained from either (2.89) or (2.91) are to be put together to the other six originated by the sp^2 block. Since the total number of valence electrons to be arranged in each cell is eight, only the first four bands are completely filled. The block sp^2 yields three bonding and three antibonding bands. Bonding bands are occupied bands at lower energies, while antibonding ones are unoccupied bands at higher energies. Similarly, the block p_z yields one bonding and one antibonding band. To conclude, three out of the first four occupied bands are originated from the sp^2 block and one from the p_z one. This situation is displayed in Fig. 2.23.

The left panel of Fig. 2.22 displays special points in the Brillouin zone, that are labeled Γ , K and M , together with the lines connecting them. Along these lines, the bands can be determined after using either (2.89) or (2.91). The result is illustrated in the left panel of Fig. 2.24. Dashed curves represent the result of the calculation from (2.91) with γ and σ fixed to the values $\gamma = -3.033$ and $\sigma = 0.129$. Solid curves represent the results from (2.89) with $\gamma = -3.033$ and $\sigma = 0$. In both cases, the qualitative behavior of Fig. 2.23 is reproduced, though quantitative discrepancies are clearly evident in such a simple two-parameter fit. The main qualitative reported feature is that fully filled bands, namely lying below ϵ_f , and empty bands touch each other at the point corresponding to K . Besides, the two approximations with and without overlap provide essentially the same result around point K .

Indeed, choose now the point with coordinates $(0, 2\pi/3b)$ among the equivalent K points displayed in the left panel of Fig. 2.22. The Taylor expansion of (2.88) around this point yields

$$f(k_x, k_y) = -ia\gamma(k_x - i\eta_y) , \quad (2.92)$$

where

$$\eta_y = k_y - \frac{2\pi}{3b} . \quad (2.93)$$

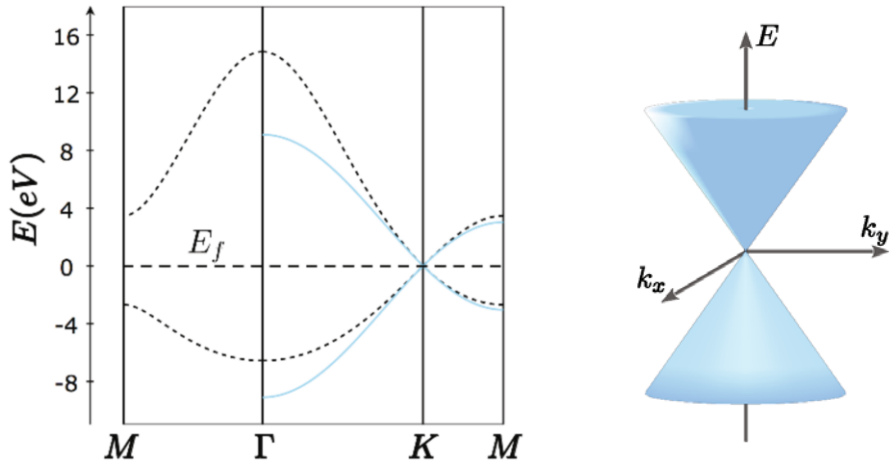


Fig. 2.24 Applications of electronic structure calculations to graphene. Left panel: band structure of graphene along the path shown in the left panel of Fig. 2.22. Filled bands are those below E_f . Dashed lines: results from (2.91). Solid lines: results from (2.89). Right panel: Sketch of the shape of the energy bands around the K point

Shifting the origin of \mathbf{k} vectors to the selected point $(0, 2\pi/3b)$, so that $\eta_y = k_y$ in (2.92), the energy bands given by (2.89) assume the form:

$$E(k_x, k_y) = E_0 \pm a\gamma\sqrt{k_x^2 + k_y^2} = E_0 \pm \frac{3}{2}a_0\gamma|k|. \quad (2.94)$$

Apart from the constant term E_0 , the bands have a conic shape with circular section, as sketched in the right panel of Fig. 2.24. The dispersion relation given by (2.94) is typical of particles with vanishing inverse effective masses, as it is also evident from the right panel of Fig. 2.24.

Carbon nanotubes. As displayed in Fig. 2.16, a slice of graphene can be cut along a given axis and rolled up to form a carbon nanotube (CNT) [10]. A carbon nanotube can be thought of as a one-dimensional system, meaning that along its axial direction a primitive cell can be singled out, whose atoms are in essence arranged on a cylinder surface. The number of atoms composing the cell depends on the way the graphene layer is wrapped up. Fig. 2.16 displays an example of primitive cell. Since CNTs are seamless objects, the corresponding wavefunctions have to obey periodic boundary conditions around the circumference. Since carbon nanotubes can be considered very long with respect to their transverse size, in fact in the range of a few microns long, periodic boundary conditions on the tube surface are appropriate and exact: the unit cell can be infinitely replicated to yield a periodic structure in the axial direction. On the other hand, if the nanotube were shorter, of the order of a few nanometers, periodic boundary conditions on the surface would still be correct, those along the axis becoming invalid instead.

A single-wall CNT is obtained by rolling up a single graphene sheet. Therefore, its microscopic structure is closely related to that of graphene and can be specified in terms of the graphene lattice vectors. As shown in Fig. 2.16, the nanotube circumference is defined by the chiral vector \mathbf{C}_h , connecting two graphene lattice sites, say A_1 and A_2 : when the atoms occupying the sites A_1 and A_2 are rolled into each other, the tube is formed. This vector can thus be expressed as a linear combination of the graphene unit vectors: $\mathbf{C}_h = n\mathbf{a}_1 + m\mathbf{a}_2$, with n and m two integers. The pairs of integers (m, n) completely identify the CNT geometry. The smallest graphene lattice vector \mathbf{T} that is perpendicular to \mathbf{C}_h , defines instead the CNT unit cell period, as displayed in Fig. 2.16. The special cases of $(n, 0)$ and (n, n) nanotubes are referred to as zigzag and armchair CNTs, according to the geometric pattern along the circumference.

Depending on the chiral vector and therefore on the (n, m) indexes, CNTs can be either metallic or semiconducting. The former case occurs when $n - m = 3l$ with l an integer, the latter when $n - m = 3l \pm 1$. The band structure and density of states for zigzag and armchair CNTs are displayed in Fig. 2.25.

2.4 Methods for band-structure calculations

2.4.1 Plane Wave method

The Tight-Binding method is suited to investigate the evolution of energy levels from the discrete states occurring in nanocrystals to their quasi-continuum distribution into bands, while the number of atoms gets very large values. Bands in turn, may result to be separated by forbidden gaps, whose size depends on the matrix elements of the crystal potential calculated between atomic orbitals. A difficulty arises due to the possible localization of atomic orbitals in the vicinity of single atoms in the crystal. As anticipated in Sec. 2.2, a crystal is expected to have energy bands with lower energy, whose corresponding Bloch states are tightly localized in the vicinity of atoms, as well as energy bands with higher energy, whose corresponding Bloch states tend to be delocalized among atoms. It is clear that in the former case TB is expected to provide accurate solutions, while in the latter the accuracy can be rapidly lost.

Whenever accurate results are required for higher energy bands, two strategies can be adopted. First, a minimal base can be added, to account for the higher energy orbitals. Second, the model can be extended to include the matrix elements between second and third neighbors. The latter strategy is often used. Whatever the choice be, the intrinsic simplicity of the original method is being spoiled, and complexity grows up, especially if the matrix elements are used as parameters to be determined after fitting experimental data. Such an increase in the number of parameters makes the optimization and fitting process hard and ambiguous.

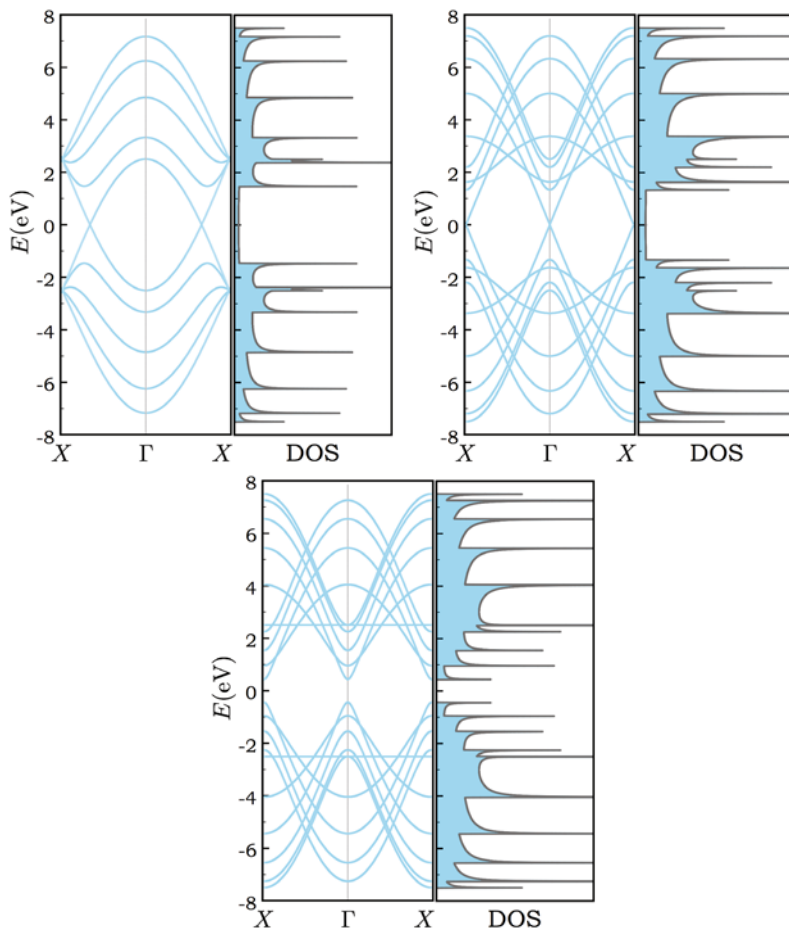


Fig. 2.25 Applications of electronic structure calculations to Carbon Nano Tubes (CNT). For each figure, the band structure calculated from Tight-Binding approximation is displayed on the left panel and the corresponding density of states on the right panel. Top left: (5,5) metallic CNT. Top right: (9,0) metallic CNT. Bottom: (10,0) semiconducting CNT

In the present section an alternative method is developed to determine the band structure of crystals, referred to as free-electron approximation. The method takes its start from the opposite consideration with respect to the Tight-Binding: the state that best approximates the real one is taken to be a linear combination of completely delocalized wavefunctions. These are of course plane waves. The main advantage of this method is its simplicity. The disadvantages mirror those of TB method: while high-energy states are expected to be well approximated, worse results come out for the lower energy states. The latter disadvantage can be overcome, and this is done introducing the concept of pseudopotential, that is the matter of Sec. 2.4.3.

2.4.1.1 Schrödinger equation based on plane waves

The Schrödinger equation for the wavefunction $\psi(\mathbf{r})$ of an electron in a crystal is

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) \right] \psi(\mathbf{r}) = E \psi(\mathbf{r}), \quad (2.95)$$

where the single-particle potential $V(\mathbf{r})$ has the periodicity of the Bravais lattice. From (1.24) it follows that the function $V(\mathbf{r})$ can be represented as a linear combination of plane waves on the Bravais lattice, namely

$$V(\mathbf{r}) = \sum_{\mathbf{G}} U(\mathbf{G}) e^{i\mathbf{G} \cdot \mathbf{r}}, \quad (2.96)$$

where the Fourier coefficients $U(\mathbf{G})$ of the linear combination are formally given by

$$U(\mathbf{G}) = \frac{1}{V_c} \int_{V_c} V(\mathbf{r}) e^{-i\mathbf{G} \cdot \mathbf{r}} d\mathbf{r}, \quad (2.97)$$

with V_c the volume of the primitive cell. Expression (2.97) asks for two remarks. First, $V(\mathbf{r})$ is a real function, so that $U(-\mathbf{G}) = U^*(\mathbf{G})$. Second, $U(0) = V_c^{-1} \int_{V_c} V(\mathbf{r}) d\mathbf{r}$ is the average value that the potential assumes in the primitive cell. Since the potential is defined except for an additive constant, the arbitrary choice $U(0) = 0$ is useful to fix a zero for the energy.

The Bloch theorem shapes the wavefunction $\psi(\mathbf{r})$ to be $u(\mathbf{r})e^{i\mathbf{k} \cdot \mathbf{r}}$ with $u(\mathbf{r})$ a function periodic on the Bravais lattice. This result can be used to write $u(\mathbf{r})$ as a superposition of plane waves $e^{i\mathbf{G} \cdot \mathbf{r}}$ and calculate the corresponding coefficients. Before proceeding though and just for pedagogical reasons, it is an interesting exercise to derive the Bloch theorem from (2.95) and (2.96). Consider that the plane waves $e^{i\mathbf{q} \cdot \mathbf{r}}$ with \mathbf{q} defined from (2.65), are an orthonormal set on the whole crystal volume where periodic boundary conditions are imposed on the wavefunction. One can thus write

$$\psi(\mathbf{r}) = \sum_{\mathbf{q}} C(\mathbf{q}) e^{i\mathbf{q} \cdot \mathbf{r}}, \quad (2.98)$$

with the vectors \mathbf{q} given by (2.65), with no restrictions on integers v_1 , v_2 , and v_3 . Substitution of (2.98) and (2.96) into (2.95) yields:

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + \sum_{\mathbf{G}} U(\mathbf{G}) e^{i\mathbf{G} \cdot \mathbf{r}} \right] \sum_{\mathbf{q}} C(\mathbf{q}) e^{i\mathbf{q} \cdot \mathbf{r}} = E \sum_{\mathbf{q}} C(\mathbf{q}) e^{i\mathbf{q} \cdot \mathbf{r}}. \quad (2.99)$$

The vectors $\mathbf{q}' = \mathbf{G} + \mathbf{q}$ in the term with the double sum, are yet given by (2.65). With this change of variables, one obtains

$$\sum_{\mathbf{G}, \mathbf{q}} U(\mathbf{G}) C(\mathbf{q}) e^{i(\mathbf{G} + \mathbf{q}) \cdot \mathbf{r}} = \sum_{\mathbf{G}, \mathbf{q}'} U(\mathbf{G}) C(\mathbf{q}' - \mathbf{G}) e^{i\mathbf{q}' \cdot \mathbf{r}}.$$

Expression (2.99) then becomes

$$\sum_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} \left[\left(\frac{\hbar^2 q^2}{2m} - E \right) C(\mathbf{q}) + \sum_{\mathbf{G}} U(\mathbf{G}) C(\mathbf{q} - \mathbf{G}) \right] = 0. \quad (2.100)$$

Use of the plane wave orthonormal conditions on the crystal volume then leads to the result:

$$\left(\frac{\hbar^2 q^2}{2m} - E \right) C(\mathbf{q}) + \sum_{\mathbf{G}} U(\mathbf{G}) C(\mathbf{q} - \mathbf{G}) = 0.$$

Now, as shown by (1.32), the vectors \mathbf{q} can be unambiguously written as $\mathbf{q} = \mathbf{k} - \mathbf{G}$ with \mathbf{k} in the Brillouin zone and \mathbf{G} a reciprocal lattice vector. Thus,

$$\left[\frac{\hbar^2 (\mathbf{k} - \mathbf{G})^2}{2m} - E \right] C(\mathbf{k} - \mathbf{G}) + \sum_{\mathbf{G}'} U(\mathbf{G}' - \mathbf{G}) C(\mathbf{k} - \mathbf{G}') = 0. \quad (2.101)$$

Expressions (2.101) represent a homogeneous set of linear equations with respect to the unknowns $C(\mathbf{k} - \mathbf{G})$. Eigenvalues E and eigenvectors $C(\mathbf{k} - \mathbf{G})$ are determined after diagonalizing the matrix

$$H_{\mathbf{G},\mathbf{G}'} = \frac{\hbar^2}{2m} (\mathbf{k} - \mathbf{G})^2 \delta_{\mathbf{G},\mathbf{G}'} + U(\mathbf{G}' - \mathbf{G}). \quad (2.102)$$

The values of \mathbf{k} are to be fixed each time inside the first Brillouin zone by means of (2.65). If N is the total number of \mathbf{G} vectors that are used in the expansion (2.96), then the matrix $H_{\mathbf{G},\mathbf{G}'}$ turns out to be of order N . After each diagonalization executed at the given value of \mathbf{k} , this yields N eigenvalues that can be ordered from the smaller to the larger. Eigenvalues are thus indexed by an integer n and a vector \mathbf{k} in the first BZ, and from now on are indicated with the notation $E_n(\mathbf{k})$. $E_n(\mathbf{k})$ represents the energy of the n -th band at each given value of \mathbf{k} . After each diagonalization, N eigenvectors $C(\mathbf{k} - \mathbf{G})$ come out, which can be used to reconstruct back the wavefunctions in (2.98), by means of a linear combination of the N values of \mathbf{G} :

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} C(\mathbf{k} - \mathbf{G}) e^{i(\mathbf{k} - \mathbf{G})\cdot\mathbf{r}}. \quad (2.103)$$

This formula is especially interesting: after factorizing the phase $e^{i\mathbf{k}\cdot\mathbf{r}}$, the remaining sum over \mathbf{G} is the Fourier expansion of a periodic function on the Bravais lattice. This leads to the wanted alternative demonstration of Bloch theorem (2.59), since

$$\begin{aligned} \psi_{\mathbf{k}}(\mathbf{r}) &= e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}), \\ u_{\mathbf{k}}(\mathbf{r}) &= \sum_{\mathbf{G}} C(\mathbf{k} - \mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}. \end{aligned} \quad (2.104)$$

In (2.104), the functions $u_{\mathbf{k}}(\mathbf{r})$ are linear combinations of plane waves with wavenumbers given by reciprocal lattice vectors.

An interesting property can be directly deduced from (2.102). Suppose that (2.102) has been solved for a given \mathbf{k} . Taking now a \mathbf{k}' outside the first BZ, in the form $\mathbf{k}' = \mathbf{k} + \mathbf{G}$, the matrix can be once more diagonalized. Since the matrix elements of the potential depend solely on the difference between vectors of the reciprocal lattice, the very same eigenvalues are obtained:

$$E_n(\mathbf{k}) = E_n(\mathbf{k} + \mathbf{G}) . \quad (2.105)$$

This expression demonstrates that $E_n(\mathbf{k})$ is a periodic function of the reciprocal lattice, precisely as it has been found by means of the Tight-Binding method applied to a one-dimensional crystal. In much the same way, (2.105) implies that

$$\psi_{\mathbf{k}}(\mathbf{r}) = \psi_{\mathbf{k}+\mathbf{G}}(\mathbf{r}) . \quad (2.106)$$

Since $U(-\mathbf{G}) = U^*(\mathbf{G})$ and given the form of $H_{\mathbf{G},\mathbf{G}'}$, after operating the transformation $\mathbf{k} \rightarrow -\mathbf{k}$ it is found that the eigenvalues $E_n(\mathbf{k})$ possess the property

$$E_n(\mathbf{k}) = E_n(-\mathbf{k}) . \quad (2.107)$$

In essence, the Plane Wave (PW) method satisfies the Bloch theorem along with the properties of periodicity and of even parity with respect to the transformation $\mathbf{k} \rightarrow -\mathbf{k}$, found while developing the Tight-Binding method.

2.4.1.2 One-dimensional crystal

As usual, understanding passes through the study of simple examples. As for the TB method, also for the Plane Wave method the case of a one-dimensional crystal is now considered. To proceed step by step, the matrix elements of the potential appearing in (2.102) are neglected in a first approach. The matrix is now diagonal: ordering the reciprocal lattice vectors from the smaller to the larger, the corresponding values on the diagonal are the band energies

$$\varepsilon_n(k) = \frac{\hbar^2}{2m} \left(k - n \frac{2\pi}{a} \right)^2 \quad \text{with } n = 0, \pm 1, \pm 2 \dots . \quad (2.108)$$

Expressions (2.108) are parabola branches, which are to be folded into the first BZ as described in Sec. 2.3.2. The bands corresponding to this empty lattice, that is where the potential matrix elements are zero, are represented in Fig. 2.26.

If the matrix elements of the potential $U(G - G') = U[2\pi(n - n')/a] = V_{n-n'}$ were negligible with respect to $\varepsilon_n - \varepsilon_{n'}$, then the bands shown in Fig. 2.26 are a reliable first approximation. Corrections can be refined, after using perturbation theory for non-degenerate states, at all k values but those in proximity of which band crossing (and thus degeneracy) occurs. As an example, for ε_0 and $\varepsilon_{\pm 1}$ this occurs at $k = \pm \frac{\pi}{a}$. These cases are better analyzed by explicitly writing the matrix (2.102):

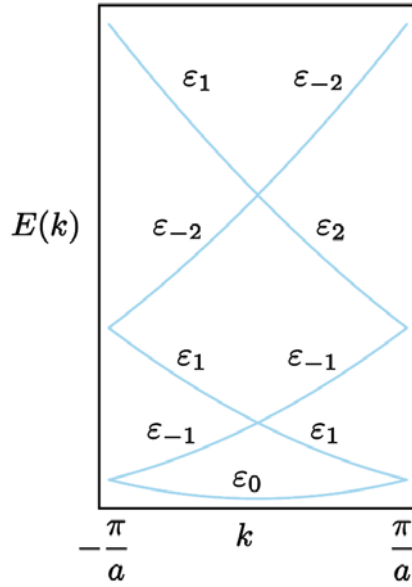


Fig. 2.26 Electronic structure with the Plane Wave method. Case of a one-dimensional crystal. Bands corresponding to an empty lattice, that is with zero potential

$$\begin{pmatrix} n, n' & 0 & 1 & -1 & 2 & -2 & \dots \\ 0 & \epsilon_0(k) - E & V_{-1} & V_1 & V_{-2} & V_2 & \dots \\ 1 & V_1 & \epsilon_1(k) - E & V_2 & V_{-1} & V_3 & \dots \\ -1 & V_{-1} & V_{-2} & \epsilon_{-1}(k) - E & V_{-3} & V_{-1} & \dots \\ 2 & V_2 & V_1 & V_3 & \epsilon_2(k) - E & V_4 & \dots \\ -2 & V_{-2} & V_{-3} & V_{-1} & V_{-4} & \epsilon_{-2}(k) - E & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (2.109)$$

It is clear that if perturbative solutions to ϵ_0 and ϵ_1 are being looked for in the vicinity of zone boundaries, where they are degenerate, one can just extract from (2.109) the individual block

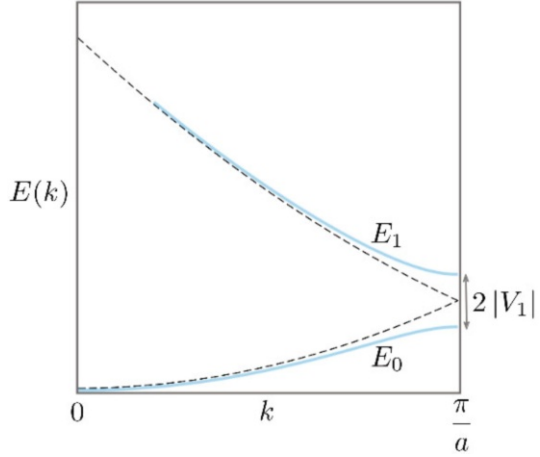
$$\begin{pmatrix} \epsilon_0(k) - E & V_{-1} \\ V_1 & \epsilon_1(k) - E \end{pmatrix}.$$

Diagonalization of this block yields the matrix elements of the perturbation in the subspace spanned by the degenerate states. The result is

$$E_{0,1}(k) = \frac{1}{2} [\epsilon_0(k) + \epsilon_1(k)] \pm \frac{1}{2} \sqrt{(\epsilon_0(k) - \epsilon_1(k))^2 + 4|V_1|^2}. \quad (2.110)$$

It is immediately seen that the two bands, which at the beginning were degenerate at $k = \pi/a$, are now split up with a gap $E_1(\pi/a) - E_0(\pi/a) = 2|V_1|$, that opens at point $k = \pi/a$. This is displayed in Fig. 2.27.

Fig. 2.27 Electronic structure with the Plane Wave method. Case of a one-dimensional crystal. Solid lines: bands calculated after inclusion of potential matrix elements within perturbation theory. A gap is opened up at the BZ boundary. For comparison, the unperturbed, free electron, band structure is also reported with dashed lines



In a similar way, $\varepsilon_1(k)$ and $\varepsilon_{-1}(k)$ are found to be degenerate at $k = 0$. The block to be considered is now

$$\begin{pmatrix} \varepsilon_1(k) - E & V_2 \\ V_{-2} & \varepsilon_{-1}(k) - E \end{pmatrix},$$

that opens up a second gap with size $2|V_2|$. The same procedure can be applied to all degeneracy points. The final result is that bands in the first Brillouin zone display the behavior represented in [Fig. 2.28 \(b\)](#).

2.4.1.3 Energy gap for a 3D crystal

Gap opening at the zone center and boundaries is a general property of crystals, not limited to one-dimensional cases, but also to two (2D) and three (3D) dimensions. In quantum physics, two degenerate isolated states are always split up as soon as an interaction between them, even at perturbative level, is switched on: this is a structural result of how quantum mechanics and related fundamental equations works. This is clearly seen looking at the 2×2 block in (2.102): for a given \mathbf{k} , one has that $\hbar^2/(2m)(\mathbf{k} - \mathbf{G}_1)^2 \sim \hbar^2/(2m)(\mathbf{k} - \mathbf{G}_2)^2$:

$$\begin{pmatrix} \frac{\hbar^2}{2m}(\mathbf{k} - \mathbf{G}_1)^2 & U(\mathbf{G}_1 - \mathbf{G}_2) \\ U(\mathbf{G}_2 - \mathbf{G}_1) & \frac{\hbar^2}{2m}(\mathbf{k} - \mathbf{G}_2)^2 \end{pmatrix}. \quad (2.111)$$

Since two reciprocal vectors are here involved, the settings $\mathbf{q} = \mathbf{k} - \mathbf{G}_1$ and $\mathbf{G} = \mathbf{G}_2 - \mathbf{G}_1$ are convenient in (2.111), so that

$$\begin{pmatrix} \frac{\hbar^2}{2m}\mathbf{q}^2 & U(-\mathbf{G}) \\ U(\mathbf{G}) & \frac{\hbar^2}{2m}(\mathbf{q} - \mathbf{G})^2 \end{pmatrix}. \quad (2.112)$$

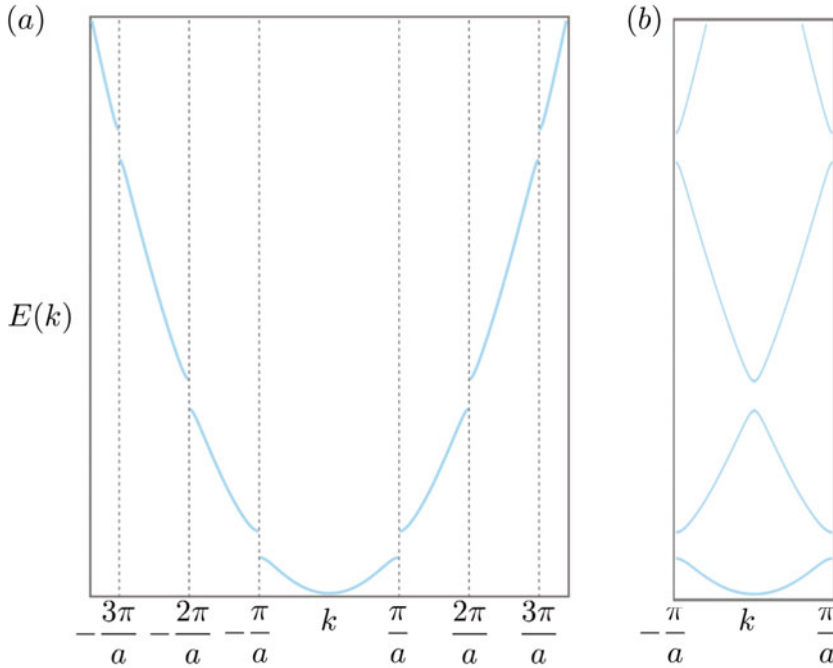


Fig. 2.28 Electronic structure with the Plane Wave method. The band structure showing gap opening at π/a is schematically displayed in unfolded (a) and folded (b) pictures. To catch the basic structure, the case of a 1D crystal is shown

Considering the above discussion and noticing that $U^*(\mathbf{G}) = U(-\mathbf{G})$, a gap opens up when $|\mathbf{q}| = |\mathbf{q} - \mathbf{G}|$. Such a circumstance takes place whenever \mathbf{q} points to a plane orthogonal to \mathbf{G} and passing through $\mathbf{G}/2$, as shown in Fig. 2.29. The plane which the vector \mathbf{q} points to is the Bragg plane, as already discussed in Chap. 1 in relation to X-ray diffraction. A special case is met when $\mathbf{G}_1 = -\mathbf{G}_2$ and the Bragg plane passes through the zone center, its equation being $\mathbf{k} \cdot \mathbf{G}_1 = 0$.

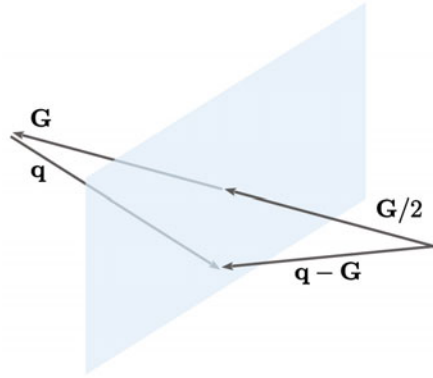
More generally,

Concept

At whatever \mathbf{k} -vector in proximity of a Bragg plane, a gap is expected to open up, with amplitude $2|U(\mathbf{G})|$.

This effect reflects on the whole band structure. To see this in a clear and simple way, one better reverts back to the one-dimensional case depicted in Fig. 2.28, where the band structure with gap openings is shown in both unfolded (a) and folded (b) pictures.

Fig. 2.29 Electronic structure with the Plane Wave method in 3D crystals. Graphical representation of the condition $|\mathbf{q}| = |\mathbf{q} - \mathbf{G}|$ identifying the direction along which the gap is expected to open up in the presence of degenerate states. The \mathbf{q} vector points to a plane perpendicular to \mathbf{G} and passing through $\mathbf{G}/2$. This is a Bragg plane



2.4.2 Wannier method

Many real systems can be close to either best TB or to best PW conditions:

Concept

Bloch functions have been constructed starting both from sets of atomic wavefunctions and from sets of plane waves. The former procedure is a base to the Tight-Binding (TB) method and is appropriate to atomic orbitals that are tightly localized to respective atoms. The latter is a base to the Plane Wave (PW) method and is convenient when electrons are as much delocalized as possible within the crystal.

In general though, many more meet conditions lying in between these two extreme limiting behaviors. Thus methods need to be developed, which are capable of effectively entering this mid region. In this section the Wannier picture is illustrated, which conceptually and, in some cases practically, may do the job.

Now, TB is easily handled as much as the orbitals are localized, the overlap between neighboring atoms being negligible. Increasing degrees of overlap progressively shifts the preference towards the PW method. The problem with TB method is that atomic orbitals on neighboring sites are not orthogonal, so that the computational complexity rapidly increases with the number of orbitals that are to be considered, in order to account for increasing delocalization of electronic wavefunctions.

Orthogonality relations (2.3) are recovered within an alternative method based on the Wannier functions $w_n(\mathbf{r} - \mathbf{R}_p)$, centered on the lattice sites \mathbf{R}_p . Indeed, a Bloch function can be exactly expressed in terms of Wannier functions according to the relation

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = \sum_p e^{i\mathbf{k} \cdot \mathbf{R}_p} w_n(\mathbf{r} - \mathbf{R}_p). \quad (2.113)$$

Multiplication of the right-hand side of (2.113) by $1 = e^{i\mathbf{k} \cdot \mathbf{r}} e^{-i\mathbf{k} \cdot \mathbf{r}}$, immediately casts (2.113) in the form of a Bloch function, that is a plane wave times a periodic function

on the Bravais lattice. Expression (2.113) is an elegant, and exact, representation of a concept that TB method implements at an approximated level:

Concept

In Wannier representation, a Bloch function is exactly assembled by taking the (Wannier) wavefunctions centered on each lattice site and connecting them by means of phases $e^{i\mathbf{k}\cdot\mathbf{R}_p}$ dictated by the Bloch wavevector \mathbf{k} and the lattice vectors \mathbf{R}_p .

Wannier functions are localized in the primitive cell and centered on a lattice site. In a crystal without base, not much difference arises between using either a TB method or a Wannier picture: the Wannier function is to first and best approximation given by an atomic orbital whenever first-neighbor overlap is negligible.

A few properties of Wannier functions may be useful. First, since Bloch functions are periodic on the reciprocal lattice, that is $\psi_{n,\mathbf{k}}(\mathbf{r}) = \psi_{n,\mathbf{k}+\mathbf{G}}(\mathbf{r})$ as in (2.106), expression (2.113) can be cast in the form

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} w_n(\mathbf{r}, \mathbf{R}) e^{i\mathbf{k}\cdot\mathbf{R}}. \quad (2.114)$$

Here,

$$w_n(\mathbf{r}, \mathbf{R}') = \frac{1}{V_r} \int_{V_r} d\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{R}'} \psi_{n,\mathbf{k}}(\mathbf{r}), \quad (2.115)$$

and V_r is as usual the BZ volume. The following property is also immediately verified:

$$w_n(\mathbf{r}, \mathbf{R}') = w_n(\mathbf{r} + \mathbf{R}, \mathbf{R}' + \mathbf{R}), \forall \mathbf{R}.$$

Thus, one may conclude that

$$w_n(\mathbf{r}, \mathbf{R}') = w_n(\mathbf{r} - \mathbf{R}', 0) = w_n(\mathbf{r} - \mathbf{R}'), \quad (2.116)$$

and

Properties

P1. By construction, Wannier functions are a complete orthonormal basis set because the Bloch functions do are.

P2. Wannier and Bloch functions are connected by Fourier transformation with respect to the pair of variables \mathbf{k} and \mathbf{R} , as in (2.115).

P3. Wannier functions depend on the difference $\mathbf{r} - \mathbf{R}$, as in (2.116).

A similar reasoning can be repeated for other periodic quantities on the reciprocal lattice. One important example are the energies $E_n(\mathbf{k})$. After Fourier transformation, one has indeed:

$$E_n(\mathbf{k}) = \sum_{\mathbf{R}} E_n(\mathbf{R}) e^{i\mathbf{k} \cdot \mathbf{R}}, \quad (2.117)$$

with

$$E_n(\mathbf{R}) = \frac{1}{V_r} \int_{V_r} d\mathbf{k} e^{-i\mathbf{k} \cdot \mathbf{R}'} E_n(\mathbf{k}). \quad (2.118)$$

The physical content of expression (2.117) is novel in the context of band structure calculations:

Concept

In Wannier picture, the band structure is assembled from the knowledge of the energies $E_n(\mathbf{R})$ at each lattice site and connected by phases $\mathbf{k} \cdot \mathbf{R}$, which the dispersion in \mathbf{k} -space thus arises from.

The material presented in this section leads to understanding that the TB and PW methods can be viewed as special realizations of the elegant Wannier picture. The conceptual elegance of Wannier picture might shatter against the issue of determining the Wannier functions. The form of Wannier functions is known for selected problems. For example, Wannier functions can be determined in a simple-cubic lattice, after using (2.115) for plane waves with wavenumbers folded in the first Brillouin zone. The result is:

$$w(\mathbf{r} - \mathbf{R}) = \frac{a^{\frac{3}{2}}}{\pi^3 \sqrt{N}} \frac{\sin(\pi X/a)}{X} \frac{\sin(\pi Y/a)}{Y} \frac{\sin(\pi Z/a)}{Z}, \quad (2.119)$$

with N the number of cells in the crystal, $\mathbf{r} - \mathbf{R} = X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k}$, and a the cube side.

As a second example, Wannier functions can be determined for a lattice potential of the form $\cos(kx) + \cos(ky) + \cos(kz)$: in particular, they are known to be related to special functions, the Mathieu functions [13]. Systems with these characteristics are currently realized in laboratory with confined quantum atomic gases cooled to temperatures of tens of nK, and surfing on optical potentials made from stationary waves of counterpropagating laser light, as discussed e.g. in [14].

2.4.3 Pseudopotential methods

Sec. 2.4.2 has clarified that an upgrade of the basis set from incomplete atomic orbitals to complete Wannier functions, though still approximated in a real calculation, conceptually and practically improves the use of Tight-Binding method. The Pseudopotential methods act as an effective upgrading of the Plane Wave method.

Aim of this section is to see how this concept works. Reverting back to the PW method, one can state that this is attractive for its simplicity and possibility of effi-

cient control of convergence in relation to the total number of plane waves included in the calculation. However, a problem arises with the convergence when the oscillatory behavior that both core and valence states manifest in the vicinity of nuclei, in order to guarantee orthogonality between themselves. For example, consider the state 1s of carbon and fix the spatial resolution Δ that is required for the given calculation. In this case, (2.103) has to include all the plane waves with vectors \mathbf{G} , whose size is in the interval in $[0, G_{\max}]$: here $G_{\max} = 2\pi/\Delta$, as stated from general properties of Fourier transforms. Once the volume V_r of the first BZ is given, the number N of needed plane waves doing the job is

$$N = \frac{4\pi}{3} G_{\max}^3 \frac{1}{V_r} = \frac{4\pi}{3} \left(\frac{2\pi}{\Delta} \right)^3 \frac{1}{V_r}. \quad (2.120)$$

Here, $V_r = (2\pi)^3/V_c$ in terms of the primitive-cell volume V_c . In order to catch the order of magnitude, consider a realistic case, for example diamond. In this case $V_c = a^3/4$, so that a simple calculation yields then

$$N = \frac{\pi}{3} \left(\frac{a}{\Delta} \right)^3. \quad (2.121)$$

In diamond, $a = 3.57 \text{ \AA} = 6.75 a_b$, in terms of Bohr radius a_b . The 1s-state wavefunction gets its first maximum value at about $0.3a_b$, so that one might reasonably want to fix the resolution at level $\Delta = 0.1a_b$. With this choice, (2.121) tells that about 322000 plane waves are needed not to miss the oscillations.

Beyond the details, this turns out to be a typical order of magnitude. The corresponding computational challenge of PW method may be treacherous. The origin of the problem is mainly connected to core states, showing the faster oscillations near the nucleus. One might thus think of getting rid somehow of the core states and keep only valence states, which after all are the real chemical-bond doers.

The concept of pseudopotential origins from this idea. To understand how it works, as usual consider first the simplest case, that of an isolated atom. Be H its Hamiltonian for the valence states

$$H\psi_v = E_v\psi_v, \quad (2.122)$$

and for the core states

$$H\psi_c = E_c\psi_c. \quad (2.123)$$

The wavefunction $\psi_v(\mathbf{r})$ for the valence state can be expressed in a different basis set, which includes the core states as follows

$$\psi_v(\mathbf{r}) = \phi_v(\mathbf{r}) + \sum_c B_c \psi_c(\mathbf{r}). \quad (2.124)$$

Imposing $\langle \psi_v | \psi_c \rangle = 0$, one has

$$B_c = - \int \psi_c^*(\mathbf{r}) \phi_v(\mathbf{r}) d\mathbf{r}. \quad (2.125)$$

Substitution of (2.124) into (2.122) leads to

$$H\phi_V + \sum_c B_c H\psi_c = E_V \left(\phi_V + \sum_c B_c \psi_c \right). \quad (2.126)$$

Use of (2.123) yields an alternative expression of (2.126) as

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V_a(\mathbf{r}) \right] \phi_V(\mathbf{r}) + \sum_c (E_V - E_c) \int \psi_c^*(\mathbf{r}') \psi_c(\mathbf{r}) \phi_V(\mathbf{r}') d\mathbf{r}' = E_V \phi_V(\mathbf{r}), \quad (2.127)$$

with V_a the atomic potential. It is now useful to define an operator $V_{ps}(\mathbf{r})$ that acts on $\phi_V(\mathbf{r})$ according to

$$V_{ps}(\mathbf{r}) \phi_V(\mathbf{r}) = V_a(\mathbf{r}) \phi_V(\mathbf{r}) + \int V_p(\mathbf{r}, \mathbf{r}') \phi_V(\mathbf{r}') d\mathbf{r}', \quad (2.128)$$

where

$$V_p(\mathbf{r}, \mathbf{r}') = \sum_c (E_V - E_c) \psi_c^*(\mathbf{r}') \psi_c(\mathbf{r}). \quad (2.129)$$

Here is how the pseudopotential concept works in its mathematical representation. First, note that the always positive difference $E_V - E_c$ has appeared in (2.129), which works to make V_p a slowly-varying function of space. Second, the first and second terms in the right hand of (2.128), which have opposite signs, tend to cancel with each other. Overall, the functions result to be amenable to expansions in a limited number of plane waves. The reason why $V_{ps}(\mathbf{r})$ is named pseudopotential is clarified as follows. With its definition, the Schrödinger equation for $\phi_V(\mathbf{r})$ becomes

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V_{ps}(\mathbf{r}) \right] \phi_V(\mathbf{r}) = E_V \phi_V(\mathbf{r}). \quad (2.130)$$

Expression (2.130) has the form of an equation for a single particle in the potential $V_{ps}(\mathbf{r})$. The whole complexity of core states is effectively embodied in the integral term (2.128). Along these lines, $V_{ps}(\mathbf{r})$ is a pseudopotential, including the sum of a real atomic potential and a term originated by the orthogonalization condition between valence and core states. Notice that V_{ps} is a first example of nonlocal potential met in the present textbook. A potential is nonlocal when its action on a wavefunction at position \mathbf{r} depends on what happens on different locations \mathbf{r}' . This is mathematically expressed as an integral over \mathbf{r}' of $V_p(\mathbf{r}, \mathbf{r}') \phi_V(\mathbf{r}')$, as in (2.128).

One may remark that the problem complexity has just been moved from one place to another. In fact, the problem is more easily handled instead. Indeed, notice that the faster oscillating part of $\psi_V(\mathbf{r})$ is concentrated in proximity of nuclei and represented by $\psi_c(\mathbf{r})$ in (2.124). Thus, $\phi_V(\mathbf{r})$ turns out to be a slowly varying function, possibly with no nodes. As a consequence, $\phi_V(\mathbf{r})$ can be represented by a moderately small number of plane waves. Alternatively, the same conclusion can be reached by adding the potential in (2.129) to the atomic potential $V_a(\mathbf{r})$ in a way to define the pseudopotential (2.128). In the core region a cancellation occurs between

$V_a(\mathbf{r})$ and the nonlocal term. This cancellation makes once more $V_{ps}(\mathbf{r})$ a slowly varying function in the vicinity of the nucleus. The possibility arises to fasten the convergence of the eigenvalues in (2.130) towards the valence eigenvalues by using a limited number of plane waves.

Fig. 2.30 illustrates a qualitative scheme accounting for the difference between atomic and pseudo-potentials, along with the difference between the exact valence wavefunction $\psi_v(\mathbf{r})$ and the pseudo wave function $\phi_v(\mathbf{r})$.

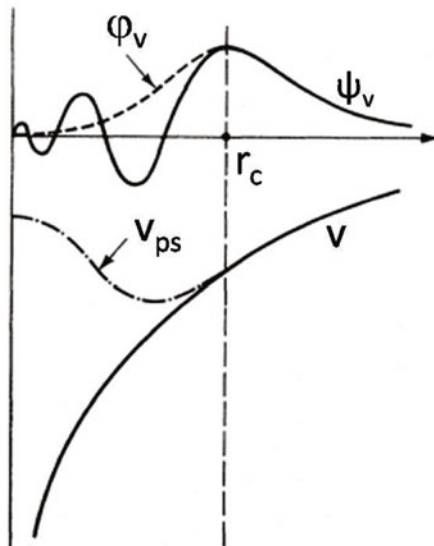


Fig. 2.30 Electronic structure with the Pseudopotential method. Schematic representation of the pseudopotential $V_{ps}(\mathbf{r})$ and the pseudo-wavefunction $\phi_v(\mathbf{r})$. For comparison, the qualitative behavior of the atomic potential $V = V_a(\mathbf{r})$ and of the exact atomic wavefunction $\psi_v(\mathbf{r})$ are sketched

The figure shows that a cut-off radius r_c can be artificially defined around the nucleus, so that pseudopotential and real potential coincide for distances larger than r_c . In the region inner to r_c , the pseudopotential is by construction a slowly varying function of space. Similarly, ϕ_v coincides with the exact wavefunction ψ_v for $r > r_c$, while it is nodeless in the inner region for $r < r_c$. This fact does the job, since a nodeless wavefunction can be represented by means of a limited number of plane waves.

From the above development it follows a further property of the pseudopotential theory, that is, $V_{ps}(\mathbf{r})$ is not univocal: different pseudopotentials can be constructed for the very same problem. This might appear a destabilizing idea. Actually, this is what makes the method powerful: as in a recipe, the possibility of varying the degree with which the various ingredients enter allows optimizations. In this case, optimization is to be intended with respect to both issues of computational convergence and of transferability between different crystalline systems.

2.4.3.1 Empirical pseudopotentials

The discussion around the TB method has enlightened the possibility of using an effective strategy to handle the many computational parameters. The main idea put forward by Slater is to use them as adjustable parameters to be fitted from experimental data, such as energy gaps, effective masses, and optical absorption coefficients might be. Precisely the same strategy is now being used to handle in practice PW and Pseudopotential methods.

In this section the pseudopotential matrix elements are handled as parameters. To see how it works, the pseudopotential is first and more usefully expressed as a sum of terms centered on individual atoms, that is

$$V_{ps}(\mathbf{r}) = \sum_{\mathbf{R}, \mu} v_{ps, \mu}(\mathbf{r} - \mathbf{R} - \mathbf{d}_{\mu}) , \quad (2.131)$$

the sum being extended to all \mathbf{R} vectors of the Bravais lattice and each atomic species labeled by μ , so that their position in the primitive cell is identified by vectors \mathbf{d}_{μ} . The matrix elements of the potential (2.131) between plane waves $e^{i\mathbf{G} \cdot \mathbf{r}}$ are now calculated:

$$V_{ps}(\mathbf{G}) = \frac{1}{V_c} \int_{V_c} V_{ps}(\mathbf{r}) e^{i\mathbf{G} \cdot \mathbf{r}} d\mathbf{r} . \quad (2.132)$$

Since $V_{ps}(\mathbf{r})$ is a periodic function, its integral over space is independent of the choice of the cell where integration is performed. Consider for example the cell with $\mathbf{R} = 0$. In this case, from (2.131) one obtains

$$V_{ps}(\mathbf{G}) = \frac{1}{V_c} \sum_{\mathbf{R}, \mu} \int_{V_c(0)} v_{ps, \mu}(\mathbf{r} - \mathbf{R} - \mathbf{d}_{\mu}) e^{i\mathbf{G} \cdot \mathbf{r}} d\mathbf{r} .$$

Executing the change of variables $\mathbf{r} - \mathbf{R} = \boldsymbol{\rho}$, the pseudopotential can be expressed as

$$V_{ps}(\mathbf{G}) = \frac{1}{V_c} \sum_{\mathbf{R}, \mu} e^{i\mathbf{G} \cdot \mathbf{R}} \int_{V_c(\mathbf{R})} v_{ps, \mu}(\boldsymbol{\rho} - \mathbf{d}_{\mu}) e^{i\mathbf{G} \cdot \boldsymbol{\rho}} d\boldsymbol{\rho} .$$

The expression of V_{ps} can be further simplified for best use. This task can be accomplished after noticing that $e^{i\mathbf{G} \cdot \mathbf{R}} = 1$, so that

$$V_{ps}(\mathbf{G}) = \frac{1}{V_c} \sum_{\mathbf{R}, \mu} \int_{V_c(\mathbf{R})} v_{ps, \mu}(\boldsymbol{\rho} - \mathbf{d}_{\mu}) e^{i\mathbf{G} \cdot \boldsymbol{\rho}} d\boldsymbol{\rho} = \frac{1}{V_c} \sum_{\mu} \int_V v_{ps, \mu}(\boldsymbol{\rho} - \mathbf{d}_{\mu}) e^{i\mathbf{G} \cdot \boldsymbol{\rho}} d\boldsymbol{\rho} ,$$

where the integral is now extended to the whole crystal volume V . A new change of variables then yields:

$$V_{ps}(\mathbf{G}) = \frac{1}{V_c} \sum_{\mu} e^{i\mathbf{G} \cdot \mathbf{d}_{\mu}} \int_V v_{ps, \mu}(\mathbf{r}) e^{i\mathbf{G} \cdot \mathbf{r}} d\mathbf{r} . \quad (2.133)$$

The primitive cell may contain atoms of different species. Thus a slightly different form of this expression would be conveniently looked for appreciate their different

contributions. The sum over μ can be performed by grouping all the contributions coming from atoms of the same species. To make the operation clear, the index μ identifying the given atomic species is associated to two different indexes: α to identify each different species and j identifying different atoms of the same species. With this indexing rearrangement, (2.133) gets the formally simpler form

$$V_{ps}(\mathbf{G}) = \sum_{\alpha} S_{\alpha}(\mathbf{G}) v_{\alpha}(\mathbf{G}) . \quad (2.134)$$

The physical content of this alternative form enlightens that pseudopotentials can be expressed as a sum of contributions coming from all the atomic species. Each contribution is in turn determined by a structural part $S_{\alpha}(\mathbf{G})$ and a form factor part $v_{\alpha}(\mathbf{G})$. The structure factor

$$S_{\alpha}(\mathbf{G}) = \sum_j e^{i\mathbf{G} \cdot \mathbf{d}_{\alpha,j}} , \quad (2.135)$$

contains information on the crystal structure. The form factor $v_{\alpha}(\mathbf{G})$ embodies information on the pseudopotential for each given atomic species in reciprocal space, obtained after Fourier transformation, that is:

$$v_{\alpha}(\mathbf{G}) = \frac{1}{V_c} \int_V v_{ps,\alpha}(\mathbf{r}) e^{i\mathbf{G} \cdot \mathbf{r}} d\mathbf{r} . \quad (2.136)$$

2.5 Many-particle effects: a first study

For simplicity the assumption has been adopted so far, that the many electrons in the crystal could be described as a system of independent particles in an effective potential. This potential should embody the effects of interactions between electrons, electrons and nuclei, and between nuclei. No specifications on shape and properties of this effective potential have been discussed so far, nor possible ways to construct it starting from the knowledge of microscopic interactions. The present section provides a first address to this topic. The pedagogical strategy oriented to build the needed knowledge is to first remain within a framework of approximations capable of seeding the main concepts and at the same time simple enough to make the problem analytically tractable. Concepts enlightened in the present section are then extended and refined in Chap. 6, leading to the elegant treatment of Density Function Theory and, in practice, to the illustration of calculations from first principles.

In the following, basic schemes and approximations are illustrated to treat one at a time the various interactions effects.

2.5.1 Interactions with ions: the Born-Oppenheimer approximation

Consider first the electron-nucleus interaction and the corresponding Hamiltonian operator describing the i -th electron with mass m and charge $-e$ at position \mathbf{r}_i , and the I -th nucleus with mass M_I and charge $Z_I e$ at position \mathbf{R}_I . In principle this is:

$$H = H_e + H_n$$

where

$$\begin{aligned} H_e &= -\frac{\hbar^2}{2m} \sum_i \nabla_i^2 + \frac{1}{2} \sum_{\substack{i,j \\ i \neq j}} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{I,i} \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} + \frac{1}{2} \sum_{\substack{I,J \\ I \neq J}} \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|}, \\ H_n &= -\sum_I \frac{\hbar^2}{2M_I} \nabla_I^2. \end{aligned} \quad (2.137)$$

Here, ∇_i^2 and ∇_I^2 indicate the corresponding Laplacian operators for particles at positions \mathbf{r}_i and \mathbf{R}_I , respectively. The full eigenvalue problem is

$$H\Psi(\bar{\mathbf{r}}, \mathbf{R}) = E\Psi(\bar{\mathbf{r}}, \mathbf{R}), \quad (2.138)$$

for the wavefunctions $\Psi(\bar{\mathbf{r}}, \mathbf{R})$ depending on position vectors of electrons $\bar{\mathbf{r}}$ and nuclei \mathbf{R} . An immediate consideration assesses that ion masses $M_I \gg m$ greatly exceed electron masses, so that the dynamical behavior of the former is correspondingly slower. The simplest and immediate approximation stems from the above consideration. In fact, within this approximation $\Psi(\bar{\mathbf{r}}, \mathbf{R})$ can be factorized as

$$\Psi(\bar{\mathbf{r}}, \mathbf{R}) = \Phi_n(\bar{\mathbf{r}}, \mathbf{R}) \chi_p(\mathbf{R}), \quad (2.139)$$

into electronic $\Phi_n(\bar{\mathbf{r}}, \mathbf{R})$ and ionic $\chi_p(\mathbf{R})$ parts. The electronic part $\Phi_n(\bar{\mathbf{r}}, \mathbf{R})$ contains the quantum numbers n being treated in a quantum framework, and it is slowly dependent on \mathbf{R} for the reason discussed above. The ionic part instead, is treated in a subsequent step. Under these conditions, the calculation can be split in two steps. First, the electron problem for Φ_n is solved:

$$H_e(\bar{\mathbf{r}}, \mathbf{R}) \Phi_n(\bar{\mathbf{r}}, \mathbf{R}) = \varepsilon_n(\mathbf{R}) \Phi_n(\bar{\mathbf{r}}, \mathbf{R}). \quad (2.140)$$

Second, the problem for $\chi_p(\mathbf{R})$ is solved:

$$[H_n + \varepsilon_n(\mathbf{R})] \chi_p(\mathbf{R}) = E \chi_p(\mathbf{R}). \quad (2.141)$$

In (2.140) and (2.141), the slowly-varying character of the electronic wavefunction with respect to \mathbf{R} is expressed by neglecting all the derivatives of $\Phi_n(\bar{\mathbf{r}}, \mathbf{R})$ with respect to \mathbf{R} . As a result,

Concept

The slowly-varying character that the heavier nuclei positions experience with respect to electronic wavefunctions, results into the possibility of solving the electronic eigenvalue problem with the nuclei positions \mathbf{R} entering just as parameters.

In this treatment, two issues come at play. First, the eigenvalues of $H_e(\mathbf{r}, \mathbf{R})$ play the role of interaction potentials for the nuclei. Second, the system total energy is given by the eigenvalues of (2.141), where indeed the eigenvalues of the electronic part act as potentials for the ions. This is consistent with the initial assumption of slower ion-to-electron dynamics. The set of equations (2.139)-(2.141) represent the so-called Born-Oppenheimer approximation to determine eigenvalues and eigenvectors of molecular, nanostructured and crystal systems.

In order to assess the validity of this approximation, it is useful to determine the order of magnitude of electronic energies as compared to those originated by ion vibrations. In a classical-physics framework, electrons can be considered tied to nuclei by means of elastic forces characterized by stiffness k . In this picture, typical oscillation frequencies ω_e of electrons in nuclei are of the order of $\omega_e = \sqrt{k/m}$ with m the electron mass. The third Newton law states that the elastic force acting on the nucleus has the same magnitude as the force of the molecule on the electron. The stiffness is thus unchanged, so that when looking at the nucleus oscillation frequency this is $\omega_n = \sqrt{k/M}$ with M the nucleus mass. One may conclude that $\omega_e/\omega_n = \sqrt{M/m} \approx E_e/E_n$, with $E_e \approx \hbar\omega_e$ and $E_n \approx \hbar\omega_n$ the electronic and nuclei vibrational energies. For nuclei with already a few nucleons and considering the proton-to-electron mass ratio, vibration energies result to be 100 times smaller than typical electronic energies. This is the reason why, in most cases, knowledge of $\varepsilon_n(\mathbf{R})$ in a lattice is needed just at a local level, namely it can be restricted to the vicinity of the equilibrium atomic positions. Local approximations of $\varepsilon_n(\mathbf{R})$ are worked out by means of the on-purpose devised mathematical tool, that is by means of Taylor series expansions around equilibrium positions. The expansions are typically worked out up to second order, which corresponds to harmonic approximation.

Born-Oppenheimer approximation is from now on kept as the reference framework in which electronic and vibrational states are determined. Most optical and transport properties of crystals and nanostructures are best predicted within this approximation. Physical situations occur though, not rarely nor even unimportant, where nuclei and electrons dynamics are so strongly entangled that the Born-Oppenheimer approximation fails. Examples of such systems are certain classes of superconductors and particularly highly-polarizable materials. These cases are referred to more advanced textbooks [12].

2.5.2 Interactions among electrons: Hartree and Hartree-Fock methods

After electron-nucleus interactions, electron-electron interactions are considered next. The problem is handled here at the simplest level, introducing the Hartree and Hartree-Fock approximations and methods. Within these methods,

Concept

The system of interacting particles is effectively mapped in terms of a system of independent quasi-particles subjected to an effective single-particle potential, appropriately constructed.

Though Hartree and Hartree-Fock approaches are simple, the concepts underlying this mapping can be extended to very general situations and become a powerful tool in Chap. 6, where the Density Functional Theory is illustrated. The Hartree and Hartree-Fock methods are here applied to understand basic properties of macroscopic crystals and nanostructures.

Given the Hamiltonian H_e , the first trial to treat the problem of many interacting electrons is historically based on a specific choice in the many-electron wavefunction of the form

$$\Phi_n = \varphi_1(q_1)\varphi_2(q_2)\dots\varphi_N(q_N), \quad (2.142)$$

the so-called Hartree form. This is a product of wavefunctions $\varphi_i(q_i)$ of individual electrons, where variables $q_i \equiv (\mathbf{r}_i, s_i)$ are introduced to label space position \mathbf{r}_i and spin s_i variables. The variables \mathbf{R} do not appear in (2.142), since ions are supposed to be arranged in a given fixed configuration. Each $\varphi_j(q)$ function in turn, can be cast as a product of an orbital times a single-particle spin function, that is $\varphi_j(q) = \psi_j(\mathbf{r})\alpha_j(s)$.

The task is to evaluate the contribution given to the total system energy by the interaction energy between electrons. To this aim, the following strategy is adopted. The mean value of energy is first calculated for the system described by H_e using the approximated expression (2.142) to the true many-electron wavefunction. Understood that this is not the energy of the real system, its best value is estimated after minimizing the result against variations of the trial wavefunction, thus of the individual orbitals. By best value here it is meant that the result of the minimization yields the energy that best corresponds to the given system (2.142), which yet represents an approximation to the real one.

Consider thus the average of H_e on the trial wavefunction (2.142):

$$\begin{aligned} E(N) \equiv \langle \Phi_n | H_e | \Phi_n \rangle &= \sum_{i=1}^N \int \psi_i^*(\mathbf{r}) \left(-\frac{\hbar^2}{2m} \nabla^2 \right) \psi_i(\mathbf{r}) d\mathbf{r} \\ &\quad - e^2 \sum_{I=1}^M Z_I \int \frac{n(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|} d\mathbf{r} + \frac{e^2}{2} \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \end{aligned}$$

$$- \frac{e^2}{2} \sum_{i=1}^N \int \frac{|\psi_i(\mathbf{r})|^2 |\psi_i(\mathbf{r}')|^2}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \quad (2.143)$$

where the sums over spin index have been worked out and $n(\mathbf{r}) = \sum_{i=1}^N |\psi_i(\mathbf{r})|^2$ represents the electron density. The physical content of (2.143) is now being explored. The first two terms on the right-hand side of (2.143) are respectively the average kinetic energy and the average potential energy of electrons in the field of nuclei. The last two terms on the right side of (2.143) represent the electrostatic energy originated by the repulsive interaction between the static charge distributions $-en(\mathbf{r})$ and $-en(\mathbf{r}')$ at two different positions \mathbf{r} and \mathbf{r}' : this is named direct correlation term, corrected by the electrostatic self-energy of the single orbitals. Notice that all these terms have a classical analogue. Also, note that the orthogonality condition $\langle \varphi_i | \varphi_j \rangle = \delta_{i,j}$ does not require orthogonality of $\psi_i(\mathbf{r})$ and $\psi_j(\mathbf{r})$ if the spin functions of φ_i and φ_j are different, while it does if they are the same.

$E(N)$ is now to be minimized against variations of each $\psi_i(\mathbf{r}_i)$ function. Along these lines, $E(N)$ is not just a function of N , it is a functional of $\psi_i(\mathbf{r}_i)$ that enters in the integrals on the right-hand side of (2.143). Minimization of this functional against variations of $\psi_i(\mathbf{r}_i)$ requires a process of differentiation and search for zeros of functional derivatives, in a way similar to what is usually done in stationary problems for real functions of real variables. Appendix 2.10 provides the basic tools to handle functional differentiation and calculus of variations, yet in the presence of additional constraints. There, a variational theorem is also demonstrated in a simple case, stating that a functional $E(N)$ that is stationary with respect to all possible variations of the orbitals yields a value for the energy of the fundamental state that is best within the structural approximations adopted for H_e and Φ_n .

Using the tools developed in Appendix 2.10, one finds that stationary conditions for $E(N)$ against variations of each $\psi_i(\mathbf{r}_i)$, together with the normalization condition $\langle \varphi_i | \varphi_j \rangle = \delta_{i,j}$, lead to the equations:

$$-\frac{\hbar^2}{2m} \nabla^2 \psi_i(\mathbf{r}) - e^2 \sum_{I=1}^M \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} \psi_i(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r}') - |\psi_i(\mathbf{r}')|^2}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \psi_i(\mathbf{r}) = E_i \psi_i(\mathbf{r}) , \quad (2.144)$$

In (2.144), E_i are parameters that, as described in Appendix 2.10 under the name of Lagrange multipliers, account for normalization constraints on $\psi_i(\mathbf{r})$. The physical content of (2.144) can be described as well after defining the function

$$V_{H,i}(\mathbf{r}) \equiv -e^2 \sum_{I=1}^M \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} + e^2 \int \frac{n(\mathbf{r}') - |\psi_i(\mathbf{r}')|^2}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' .$$

It is thus seen that the set of equations (2.144) is formally equivalent to a set of N Schrödinger equations for the particle wave functions $\psi_i(\mathbf{r})$ with $i=1, \dots, N$, moving in an effective potential $V_{H,i}(\mathbf{r})$ that self-consistently depends on $\psi_i(\mathbf{r})$ and the charge density $n(\mathbf{r})$. Along these lines, the Hartree method is also known as self-consistent field method. The solutions of (2.144) are indeed to be found by iterative methods.

The above procedure is named Hartree method. It is based on two substantial approximations, besides those already discussed and leading to the form H_e for the electron Hamiltonian. First, the trial wave function (2.142) assumes that one-particle orbitals are reliable representations of individual electron orbitals in the crystal. Second, it completely misses the Pauli principle, since the product of the single-particle orbitals is not antisymmetrized. The former approximation may fail for example in strongly interacting electron systems, where the form of individual electron orbitals is significantly correlated to that of the other electrons in the system. However, elegant ways out can be devised as illustrated in Chap. 6, which frame the conceptual simplicity of the single-particle approach, extending it to even strongly interacting conditions.

The latter approximation is instead qualitatively structural. A first obvious step to overcome the Hartree limitation is to antisymmetrize the wavefunction (2.142). To this aim, single-particle orbitals are usefully arranged in a matrix and the following so-called Slater determinant is calculated:

$$\Phi(q_1, q_2, \dots, q_N) = \frac{1}{\sqrt{N!}} \text{Det} \begin{pmatrix} \varphi_1(q_1) & \varphi_1(q_2) & \cdots & \varphi_1(q_N) \\ \varphi_2(q_1) & \varphi_2(q_2) & \cdots & \varphi_2(q_N) \\ \vdots & \vdots & & \vdots \\ \varphi_N(q_1) & \varphi_N(q_2) & \cdots & \varphi_N(q_N) \end{pmatrix}, \quad (2.145)$$

The matrix is arranged in a way that the column index spans position and spin variables for each fixed state, while the row index spans the states for each fixed position and spin. The resulting determinant is known as Hartree-Fock trial wave function, in terms of orbitals $\varphi_i(q)$ that are yet normalized and orthogonal in pairs. Use of Slater determinant guarantees antisymmetrization of the wavefunction after exchange of two particles. Two are the possible operations doing the job: either swap positions and spins between two particles $q_i \leftrightarrow q_j$, corresponding to the exchange of two columns in the matrix, or swap the wavefunctions of two particles $\varphi_i(q_k) \leftrightarrow \varphi_j(q_k) \forall k = 1, 2, \dots, N$, corresponding to the exchange of two rows in the matrix. In any event, exchange of two columns or rows in a matrix for an odd number of times swaps the sign of the determinant, a property from algebra that represents here the physical concept of Pauli principle.

As for the Hartree approximation, the functional $E(N)$ is now to be calculated and minimized. Appendix 2.11 details the procedure to compute $E(N) = \langle \Phi | H_e | \Phi \rangle$. One finds:

$$\begin{aligned} E(N) = \langle \Phi | H_e | \Phi \rangle &= \sum_{i=1}^N \int \psi_i^*(\mathbf{r}) \left(-\frac{\hbar^2}{2m} \nabla^2 \right) \psi_i(\mathbf{r}) d\mathbf{r} \\ &- e^2 \sum_{I=1}^M Z_I \int \frac{n(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|} d\mathbf{r} + \frac{e^2}{2} \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \\ &- \frac{e^2}{2} \sum_{i,j=1}^N \delta_{m_i, m_j} \int \frac{\psi_i^*(\mathbf{r}) \psi_i(\mathbf{r}') \psi_j(\mathbf{r}) \psi_j^*(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}'. \end{aligned} \quad (2.146)$$

The first two terms at the right-hand side of (2.146) are the same as those appearing in the Hartree formulation (2.143); furthermore if the contribution $i = j$ of the fourth term is added to the third one, the electrostatic contribution in (2.143) is recovered. Therefore the terms $i \neq j$ in the last contribution at the right side are coming out from the required antisymmetrization of the wavefunction. These terms have no classical interpretation and are called exchange correlation terms. Along these lines the exchange term is expected to act only between electrons with same spin, as reflected by the delta function δ_{m_i, m_j} . The exchange term is also expected to lower the total energy, since the Pauli principle tends to keep electrons far apart, so that the electrostatic energy is lowered. This is reflected by the negative sign in front of the exchange term.

The stationary conditions are now looked for $E(N)$. Use of variational calculation tools now yields:

$$\begin{aligned} -\frac{\hbar^2}{2m}\nabla^2\psi_i(\mathbf{r}) - e^2\sum_{I=1}^M\frac{Z_I}{|\mathbf{r}-\mathbf{R}_I|}\psi_i(\mathbf{r}) + e^2\int\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}'\psi_i(\mathbf{r}) \\ - e^2\sum_{j=1}^N\delta_{m_i, m_j}\int\frac{\psi_i(\mathbf{r}')\psi_j(\mathbf{r})\psi_j^*(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' = E_i\psi_i(\mathbf{r}). \end{aligned} \quad (2.147)$$

The set of equations (2.147) is again formally equivalent to a set of N Schrödinger equations for the particle wave functions $\psi_i(\mathbf{r})$ with $i = 1, \dots, N$ moving in a self-consistent effective potential. This time, the effective Hartree-Fock potential is non-local due to the form of the exchange Fock term. The latter is a striking consequence of the antisymmetrization procedure embodied in the Slater determinant. A subtle but important remark here is that the structure of the Slater determinant in essence fixes the positions of the nodes of the overall fermionic wavefunction. Besides questioning the choice of free atomic orbitals in the Hartree approximation, strong interactions within the system can in principle change the actual positions of the nodes. When this is the case, refined approximations are needed for the nodal wavefunction structure.

While procedures to treat the Fock term and solve the equation sets (2.144) and (2.147) are illustrated in Chap. 6, the question is now answered on the physical meaning of the relationship, if any, between the total energy $E(N)$ and the energies of the single particles E_i , as they are worked out through the self-consistent solution of either (2.144) or (2.147). To this aim, (2.147) can be multiplied by $\psi_i^*(\mathbf{r})$ and then integrated over \mathbf{r} while summing over index i . One then obtains:

$$\begin{aligned} \sum_{i=1}^N E_i = E_N + \frac{e^2}{2}\int\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' \\ - \frac{e^2}{2}\sum_{i,j=1}^N\delta_{m_i, m_j}\int\frac{\psi_i^*(\mathbf{r})\psi_i(\mathbf{r}')\psi_j(\mathbf{r})\psi_j^*(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}'. \end{aligned} \quad (2.148)$$

Expression (2.148) shows that the sum of all the single-particle energies exceeds the total energy, since electrostatic and exchange energies have been counted twice.

The same result is found if one starts from (2.144) and then compares with (2.143), except for the absence of the exchange term.

Given that (2.144) and (2.147) are formally equivalent to single-particle Schrödinger equations in a self-consistent field due to ions and the other electrons, the meaning of E_i is found by the following procedure, connecting E_i to the concepts of affinity and ionization.

Procedure

Step 1. Consider $N - 1$ electrons and take the trial wavefunctions (2.142) as a product of $N - 1$ orbitals with elimination of the j -th, or else the trial wavefunction (2.145) where the j -th either row or column is eliminated.

Step 2. If at the end of this new self-consistency process, the newly found orbitals are not significantly different from those found for the N -electron problem, then the following result is obtained:

$$E_j = E(N) - E(N - 1) . \quad (2.149)$$

If this is the case, then E_j represents the ionization energy for the j -th electron.

Step 3. In similar manner, if the wavefunction (2.142) is the product of $N + 1$ orbitals or it is the determinant of a new matrix where a new row and column are added, one would find:

$$E_k = E(N + 1) - E(N) . \quad (2.150)$$

In this case, E_k represents the affinity for the electron in $\varphi_k(q_k)$ orbital.

Equations (2.149) and (2.150) express the Koopman theorem. In essence,

Concept

The single-particle energies in self-consistent Hartree-Fock calculations are the difference in total energy between systems differing by one particle: thus, they represent either electron ionization or affinity energies, depending on whether the electron is extracted or added to the system.

This special relationship between single-particle energy and total energy is discussed below for a few simple and relevant models.

2.5.3 Electron affinity and ionization energies

Electrostatic and exchange interaction terms involve pairs of electrons. The simplest model to describe these effects is by means of a unique parameter U_0 . The

model is constructed as follows. Consider a single-particle potential. For one atom, this includes nucleus and core-electron parts, along with their own correlation and exchange terms. Each atom in a crystal or nanostructure contributes with the same terms. Valence electrons contribute through electrostatic and exchange terms.

Examples

Here follow a few examples of calculation of electron and ionization energies, that are also useful to introduce a significant concept on simple grounds.

1. Consider a system with only one electronic degenerate energy level ε , so that the total energy of all the N valence electrons results to be:

$$E(N) = N\varepsilon + \frac{U_0}{2}N(N-1) , \quad (2.151)$$

with U_0 the parameter embodying all the energy contributions for each pair of valence electrons.

The chemical potential is defined as the rapidity with which the energy of the system changes after either adding one particle to or subtracting it from the system. In mathematical form:

$$\mu = \frac{\partial E_N}{\partial N} . \quad (2.152)$$

In the case under study, μ results to be

$$\mu = \varepsilon + U_0N - \frac{U_0}{2} . \quad (2.153)$$

If N_0 is the number of electrons of the neutral atom or solid, electron affinity and ionization energies result to be:

$$A = E(N_0 + 1) - E(N_0) = \varepsilon + U_0N_0 , \quad (2.154)$$

and

$$E_{ion} = E(N_0) - E(N_0 - 1) = \varepsilon + U_0(N_0 - 1) . \quad (2.155)$$

Expressions (2.153)-(2.155) lead to $\mu = (A + E_{ion})/2$. These expressions tell that the processes of adding and subtracting one electron require different amounts of energy. The two energies A and E_{ion} differ by μ as expressed by (2.153). In the limiting case when $N_0 \gg 1$, μ , A , and E_{ion} do coincide, as expected.

2. Consider now a two-level system: the lowest with energy ε_1 and the highest with ε_2 . This time, interaction terms are related only to transitions from level ε_1 to ε_2 . Assume that each electron has a spin and that each energy state, ε_1 and ε_2 , can be occupied with no more than two electrons. Be $n_\sigma^{(1)}$ and $n_\sigma^{(2)}$ the number of electrons with spin σ occupying levels with energies ε_1 and ε_2 , respectively. The total energy is

$$\begin{aligned}
E(N) = & \varepsilon_1 \sum_{\sigma} n_{\sigma}^{(1)} + \varepsilon_2 \sum_{\sigma} n_{\sigma}^{(2)} \\
& + U \sum_{\sigma} \left[n_{\sigma}^{(1)} n_{-\sigma}^{(1)} + n_{\sigma}^{(2)} n_{-\sigma}^{(2)} \right] - 2\varepsilon_0 \sum_{\sigma} n_{\sigma}^{(1)} n_{\sigma}^{(2)}. \quad (2.156)
\end{aligned}$$

The first two terms on the right-hand side correspond to the single-particle energy in terms of occupation numbers $n_{\sigma}^{(1)}$ and $n_{\sigma}^{(2)}$. The third term represents the repulsion between two electrons posed in the same state ($U > 0$). The fourth represents the attractive interaction with $\varepsilon_0 > 0$, between two electrons with same spin on different states: that is the exchange term. The total energy of the system in (2.156) can be calculated after assigning each time the given value of N . The following cases emerge, each characterized by a different enumeration of the states:

- $N = 1$. Two states with energy ε_1 , and two with energy ε_2 . Promoting one electron from level 1 to level 2 requires energy $\varepsilon_2 - \varepsilon_1$
- $N = 2$. One state with two electrons on level 1 and energy $E_1(2) = 2\varepsilon_1 + 2U$. A second state is obtained after promoting one electron on level 2 and leaving its spin unchanged, so that its energy $E_2(2)$ is:

$$\begin{aligned}
E_2(2) = & \varepsilon_1 + \varepsilon_2 = 2\varepsilon_1 + 2U + \varepsilon_2 - \varepsilon_1 - 2U \\
= & E_1(2) + (\varepsilon_2 - \varepsilon_1 - 2U). \quad (2.157)
\end{aligned}$$

A third state is obtained by promoting again an electron on level 2 after flipping its spin, so that the state energy is:

$$\begin{aligned}
E_3(2) = & \varepsilon_1 + \varepsilon_2 - 2\varepsilon_0 = 2\varepsilon_1 + 2U + \varepsilon_2 - \varepsilon_1 - 2U - 2\varepsilon_0 \\
= & E_1(2) + (\varepsilon_2 - \varepsilon_1 - 2U - 2\varepsilon_0). \quad (2.158)
\end{aligned}$$

A fourth state is obtained by promoting both electrons on level 2, corresponding to the energy

$$E_4(2) = 2\varepsilon_2 + 2U = E_1(2) + 2(\varepsilon_2 - \varepsilon_1). \quad (2.159)$$

Compared to the state in which the two particles have energy ε_1 , states with $E_2(2)$, $E_3(2)$, and $E_4(2)$ require energies respectively equal to $\varepsilon_2 - \varepsilon_1 - 2U$, $\varepsilon_2 - \varepsilon_1 - 2U - 2\varepsilon_0$, and $2(\varepsilon_2 - \varepsilon_1)$. Except for the case with $E_4(2)$, when two particles are promoted, the cases with $E_2(2)$ and $E_3(2)$ require single-particle promotion energies smaller than $\varepsilon_2 - \varepsilon_1$.

- $N = 3$. States with three electrons are obtained from the state with four electrons (see below), by subtracting one electron from either level 1 or level 2. One obtains two pairs of states with energies

$$\begin{aligned}
E_1(3) = & \varepsilon_1 + 2\varepsilon_2 + 2U - 2\varepsilon_0 = E(4) - \varepsilon_1 - 2U + 2\varepsilon_0, \\
E_2(3) = & 2\varepsilon_1 + \varepsilon_2 + 2U - 2\varepsilon_0 = E(4) - \varepsilon_2 - 2U + 2\varepsilon_0. \quad (2.160)
\end{aligned}$$

The difference is $E_2(3) - E_1(3) = -(\epsilon_2 - \epsilon_1)$, as in the $N = 1$ case but the sign. These types of states are called hole states, that is states corresponding to the absence of one electron in a completely filled band. This concept is systematically used in Chap. 5.

- $N = 4$. The state with four electrons is univocal and has energy

$$E(4) = 2\epsilon_1 + 2\epsilon_2 + 4U - 4\epsilon_0. \quad (2.161)$$

This last example illustrates a few significant characteristics of states with many particles.

Concept

States with $N = 1$ (single-electron states) and with $N = 3$ (single-hole states) are quite similar, since they behave like single-particle states. States with $N = 2$ are peculiar, since they imply an energy for the promotion of one particle from a lower to a higher energy state, that is smaller than $\epsilon_2 - \epsilon_1$.

This peculiar aspect is deepened in Sec. 2.5.6 where so-called excitonic effects are discussed.

Quick Questions

Q11. Is it legitimate to consider levels with $N = 2$ as levels with two holes?

Answer. Yes. The energies of the four levels can be expressed as:

$$(a) E_1(2) = 2\epsilon_1 + 2U = E(4) - 2\epsilon_2 - 2U + 4\epsilon_0;$$

$$(b) E_2(2) = \epsilon_1 + \epsilon_2 = E(4) - \epsilon_1 - \epsilon_2 - 4U + 4\epsilon_0;$$

$$(c) E_3(2) = \epsilon_1 + \epsilon_2 - 2\epsilon_0 = E(4) - \epsilon_1 - \epsilon_2 - 4U + 2\epsilon_0;$$

$$(d) E_4(2) = 2\epsilon_2 + 2U = E(4) - 2\epsilon_1 - 4U + 4\epsilon_0.$$

One then notices that the energies needed to promote electrons coincide with those calculated above.

2.5.4 Determination of electronic spectrum in the field of nuclei: jellium model

A simple application of the Hartree-Fock method is in systems where ionic charges can be considered to good approximation as uniformly distributed in the solid. In this case, the plane waves $\psi_i(\mathbf{r}) = V^{-1/2}e^{i\mathbf{k}_i \cdot \mathbf{r}}$ are solutions to equations (2.147): due to system neutrality, the two Coulomb potentials originated by the uniform ionic charge distributions cancel with each other. Kinetic and exchange terms do remain. Given \mathbf{k} , the kinetic term is immediately calculated after inserting this form

of $\psi_i(\mathbf{r})$ into (2.147) and yields $\hbar^2 k^2 / (2m)$. The exchange term is calculated in Appendix 2.12. Overall, the eigenvalues are given by:

$$E(k) = \frac{\hbar^2 k^2}{2m} - \frac{2e^2 k_f}{\pi} g\left(\frac{k}{k_f}\right),$$

$$g(x) = \frac{1}{2} + \frac{1-x^2}{4x} \ln \frac{|1+x|}{|1-x|}, \quad (2.162)$$

with k_f the Fermi wavevector. The following properties of these eigenvalues deserve a comment:

Properties

P1. In the long wavelength limit $k \rightarrow 0$, one has $E(k) \simeq \hbar^2 k^2 / (2m^*) + E_0$ with $E_0 = -2\pi e^2 k_f / \pi$ and $1/m^* = 1/m + 4e^2 / (3\hbar^2 \pi k_f)$. The energy spectrum is modified by the constant term, accounting for electronic exchange interactions and by the presence of an effective electron mass $m^* < m$. Introducing the concept of average particle distance r_0 in terms of electron density n by means of $4\pi r_0^3 n / 3 = 1$, E_0 can be expressed in the form $E_0 = -(18/\pi^2)^{1/3} e^2 / r_0$.

P2. In the limiting case $k \rightarrow \infty$, one has $E(k) \simeq \hbar^2 k^2 / (2m)$, namely the electron behaves as if it were a free particle. This is indeed and from now on referred to as the free-particle limit.

P3. The function $E(k)$ does have continuity properties at $k = k_f$, whereas its derivative does not: $|(\partial E(k))(\partial k)|_{k_f} = \infty$. Since the derivative of energy with respect to k is related to the electron group velocity, thus electron group velocities at k_f lose any physical meaning.

In real calculations, the nonlocal exchange term in (2.147) is usually approximated by a local one according to the following procedure, due to Slater:

Procedure

Step 1. Substitute the exchange term in (2.147) with

$$V_{xc}(\mathbf{r}) = - \frac{\sum_{i,j} e^2 \int d\mathbf{r}' \psi_j^*(\mathbf{r}') \psi_i(\mathbf{r}') \frac{1}{|\mathbf{r}-\mathbf{r}'|} \psi_i^*(\mathbf{r}) \psi_j(\mathbf{r})}{\sum_i \psi_i^*(\mathbf{r}) \psi_i(\mathbf{r})}. \quad (2.163)$$

Step 2. Calculate $V_{xc}(\mathbf{r})$ after using the plane waves $\psi_i(\mathbf{r}) = V^{-1/2} e^{i\mathbf{k}_i \cdot \mathbf{r}}$, with V the system volume. One then obtains: $V_{xc}(\mathbf{k}_i) = -\sum_j V^{-1} 4\pi e^2 / |\mathbf{k}_i - \mathbf{k}_j|^2$. This is detailed in Appendix 2.12.

Step 3. Substitute the sum with an integral between 0 and k_f , fixing k_f from the density through $k_f^3 = 3\pi^2 n$. As detailed in Appendix 2.12, this yields

$$V_{xc}(k_i) = -2 \frac{e^2}{\pi} k_f g(x)$$

where $g(x)$ defined as in (2.162) is a function decreasing from 1 to 0 in the interval $0 \leq x \leq 1$.

Step 4. Calculate the average value of $g(x)$ by means of

$$\frac{\int_0^1 dx x^2 g(x)}{\int_0^1 x^2 dx} = \frac{3}{4}.$$

Step 5. The final results is:

$$V_{xc} = -3e^2 \left(\frac{3}{8\pi} n \right)^{\frac{1}{3}}. \quad (2.164)$$

Note that (2.164) has been used in Chap. 1 to treat metallic bonding. Different approximations are elaborated in Chap. 6.

2.5.5 Determination of electronic bands of metals, semiconductors and insulators

Equation (2.147) along with appropriate approximations for the exchange term has been used during a few decades to determine electronic energy bands of metals, semiconductors and insulators. Results are usually nicely compared to experimental data. Three examples are illustrated in the following: one for a metallic compound, Copper within a pseudopotential method, and two for semiconductor compounds, Si and GaAs by a semi-empirical pseudopotential method, covered in detail. The Density Functional Method illustrated in Chap. 6 comes to the same results in straightforward manner and with no need of empirical pseudopotentials. Textbooks can be found [4, 11] where energy bands of metals or semiconductors, other than Si and GaAs, are discussed in exhaustive and effective way. No further calculations are thus worked out here on this topic, addressing to those textbooks the student who would like to dig in.

Metals. Electronic bands of metallic Copper is for example reported in Fig. 2.31. Copper's configuration is $[\text{Ar}]3d^{10}4s^1$: in this case, at least six bands are required to account for the eleven electrons, and actually turn out to be enough. Five bands turn

out to be quite narrow between 2 and 5 eV below ϵ_f and could be referred to as d bands. The remaining levels as s bands, lying between 7 and 9 eV below ϵ_f . In fact, all the six bands are close to each other at selected \mathbf{k} vectors, so that the distinction between d and s bands is meaningless.

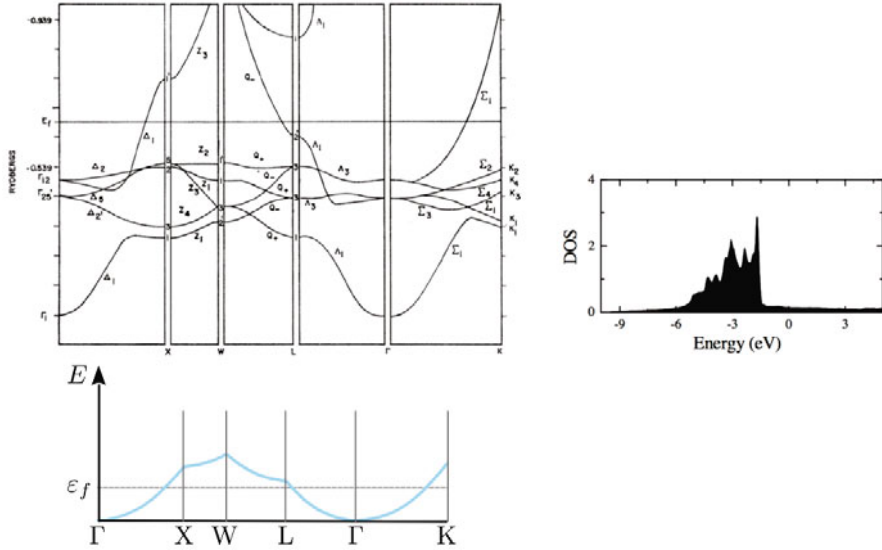


Fig. 2.31 Band structure of metals. The band structure of Copper. (Left top) Calculation from pseudopotential method. (Left bottom) Lowest band on a different scale. (Right) Electron Density of States. Reprinted with permission from [15]. Copyright (1963) by the American Physical Society

Semiconductors. Sec. 2.4.3 has introduced the semi-empirical pseudopotential method. This is here exploited to determine the band structure of two semiconductors that are relevant to applications, such as Si and GaAs. In Si the valence band maximum and conduction band minimum occur at different points in the Brillouin zone, a situation referred to as indirect gap. In GaAs valence band minimum and conduction band maximum are at the same BZ point, so that the gap is direct.

Both Si and GaAs crystallize in face-centered cubic structures with two atoms per cell, from now on labeled A and B. With reference to expressions (2.133)-(2.136), the pseudopotential can be written as:

$$V_{ps}(\mathbf{G}) = \mathbf{e}^{i\mathbf{G}\cdot\mathbf{d}} v_A(\mathbf{G}) + \mathbf{e}^{-i\mathbf{G}\cdot\mathbf{d}} v_B(\mathbf{G}), \quad (2.165)$$

where

$$v_{A(B)}(\mathbf{G}) = \frac{1}{V_c} \int_V v_{A(B)}(\mathbf{r}) \mathbf{e}^{i\mathbf{G}\cdot\mathbf{r}} d\mathbf{r}$$

and $\mathbf{d} = (1/8, 1/8, 1/8)a$. The following transformation is conveniently introduced:

$$v_A(\mathbf{G}) = \frac{1}{2} [V_S(\mathbf{G}) + V_A(\mathbf{G})]$$

$$v_B(\mathbf{G}) = \frac{1}{2} [V_S(\mathbf{G}) - V_A(\mathbf{G})] , \quad (2.166)$$

leading to

$$\begin{aligned} V_S(\mathbf{G}) &= v_A(\mathbf{G}) + v_B(\mathbf{G}) \\ V_A(\mathbf{G}) &= v_A(\mathbf{G}) - v_B(\mathbf{G}) . \end{aligned} \quad (2.167)$$

The Fourier components $V_S(\mathbf{G})$ and $V_A(\mathbf{G})$ are named symmetric and antisymmetric form factor, respectively. Substitution of (2.166) into (2.165) yields:

$$V_{ps}(\mathbf{G}) = V_S(\mathbf{G}) \cos(\mathbf{G} \cdot \mathbf{d}) + iV_A(\mathbf{G}) \sin(\mathbf{G} \cdot \mathbf{d}) . \quad (2.168)$$

Bonding in Si is qualitatively non ionic, so that $V_A(\mathbf{G}) = 0$ can be assumed, leaving only $V_S(\mathbf{G})$ to be determined. Based on the concept illustrated in Sec. 2.4.3, a small number of Fourier components can be expected to be significant while expanding the pseudopotential. In fact, all the $V_S(\mathbf{G})$ can be set to zero after a chosen threshold of $|\mathbf{G}|$, and the remaining one used as empirical parameters.

The first \mathbf{G} vectors in a fcc in units of $4\pi/a$ are given by:

$$\begin{aligned} \mathbf{G}_0 &= (0, 0, 0), \\ \mathbf{G}_3 &= (1, 1, 1), (-1, 1, 1), \dots, (-1, -1, -1), \\ \mathbf{G}_4 &= (0, 0, 2), (0, 0, -2), \dots, (-2, 0, 0), \\ \mathbf{G}_8 &= (2, 2, 0), (2, -2, 0), \dots, (0, -2, -2), \\ \mathbf{G}_{11} &= (3, 1, 1), (-3, 1, 1), \dots, (-3, -1, -1), \end{aligned} \quad (2.169)$$

where subscripts label $|\mathbf{G}|^2$. In practice, one executes the fit procedure in several steps, each time increasing the number of Fourier components, and then selects a posteriori the case yielding the best matching with experimental data on gap, effective masses and optical absorption. Best fits are usually obtained after limiting the \mathbf{G} vectors to those whose moduli are as small as $|\mathbf{G}|^2 \leq 11$. By inspection, these are the \mathbf{G} vectors listed in (2.169).

Silicon. Substituting these \mathbf{G} in (2.168) and reminding that $V_A(\mathbf{G}) = 0$, one finds that only three form factors are sufficient for Si. In conclusion, if the cut to \mathbf{G}_{11} were reasonable, all the experimental data should be reliably reproduced with only three parameters. This is indeed the case.

Before proceeding in determining the band structure though, the comparison is anticipated between these empirical form factors $V_S(\mathbf{G})$ and those determined by means of the more refined Density-Functional Theory (DFT) developed in Chap. 6. The comparison is made evident in Fig. 2.32 [16], showing that the results of the empirical calculation, represented by dots, are in excellent agreement with those worked out by DFT, represented by the solid line. Notice that in the powerful DFT technique the data are continuous, being calculated on a very dense grid of \mathbf{G} -vectors. This excellent agreement a posteriori justifies the use of a few pseudopotential matrix elements. Small q values in reciprocal space correspond to large

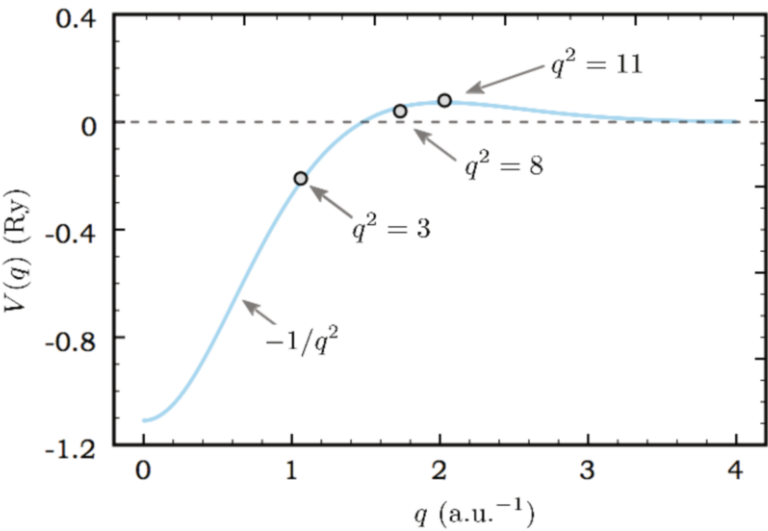


Fig. 2.32 Band structure of Si from semi-empirical pseudopotential method. Form factor as a function of the magnitude q of the \mathbf{G} vectors used, expressed in atomic units, i.e. a_b^{-1} . The comparison is shown between semi-empirical (dots) and DFT (solid line) results [16]

distances in real space, so that the form factor region close to $q^2 = 3$ in Fig. 2.32 is mainly dominated by the coulombian tail at large r values. Positive form factors at $q^2 = 8$ and 11 represent repulsive components of the pseudopotential, whose origin can be traced back to orthogonalization conditions as discussed in Sec. 2.4.3.

The values of $V_S(|\mathbf{G}|)$ as fitted from experimental data are shown in Table 2.1. The band structure of Si resulting after use of these fitted parameters is displayed in Fig. 2.33.

A few important remarks are in order:

1. Filled bands are actually in the number of four. A valence band exists as well, that has very low energy and does not appear in the figure
2. The highest valence band is twofold degenerate
3. At the BZ center (Γ) the valence band maximum is threefold degenerate
4. The first conduction band at point Γ is threefold degenerate

Table 2.1 Band structure of Si from semi-empirical pseudopotential method. Values of V_S (Ry) that fit the experimental data

| $V_S(\mathbf{G}_3)$ | $V_S(\mathbf{G}_8)$ | $V_S(\mathbf{G}_{11})$ |
|-----------------------|-----------------------|--------------------------|
| -0.21 | 0.04 | 0.08 |

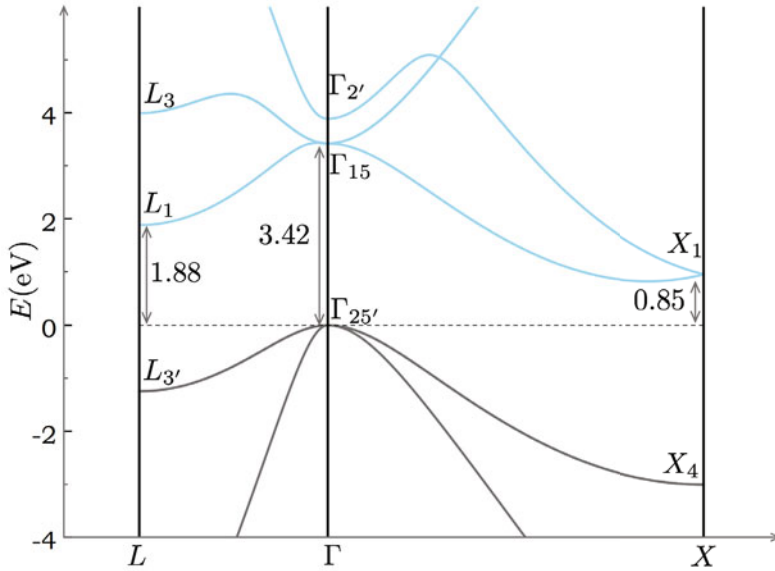


Fig. 2.33 Band structure of Si from semi-empirical pseudopotential method. Labels for the symmetry points and corresponding directions in the first BZ are those depicted in Fig. 1.29

5. The band structure is indirect, in the sense already discussed: valence band maximum is at one BZ point, $\mathbf{k} = \mathbf{0}$, while conduction band minimum is at a different point, that is $\mathbf{k} \equiv \mathbf{k}_0 = (2\pi/a)(0.85, 0, 0)$
6. Expansion of $E(\mathbf{k})$ in the vicinity of the minimum at $\mathbf{k} = (2\pi/a)(0.85, 0, 0)$ yields

$$E(\mathbf{k}) = E_g + \frac{\hbar^2}{2} \left[\frac{1}{m_l} (k_x - k_0)^2 + \frac{1}{m_t} (k_y^2 + k_z^2) \right], \quad (2.170)$$

where the effective masses are $m_t = 0.19m_e$ and $m_l = 0.916m_e$, where m_e is the free electron mass, and the gap $E_g = 0.8$ eV. For comparison, 1.1 eV is the experimental value of the gap at low temperatures

7. Since the valence band maximum is degenerate, expansions about it at Γ needs to be executed on a matrix structure. The final result is:

$$E_{\pm} = Ak^2 \pm \sqrt{B^2k^4 + C^2(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)}, \quad (2.171)$$

where $k^2 = k_x^2 + k_y^2 + k_z^2$ and A , B and C are suited coefficients. The derivation of (2.171) is not explicitly worked out. Here reported are the effective masses along special directions:

- **Direction [111].** One has $k^2 = 3k_x^2$ and thus

$$E_{\pm} = 3Ak_x^2 \pm \sqrt{9B^2k_x^4 + 3C^2k_x^4} = 3k_x^2 \left(A \pm B\sqrt{1 + \frac{C^2}{3B^2}} \right), \quad (2.172)$$

yielding the effective masses:

$$\begin{aligned} \frac{1}{m_+} &= \frac{2}{\hbar^2} \left(A + B\sqrt{1 + \frac{C^2}{3B^2}} \right) \\ \frac{1}{m_-} &= \frac{2}{\hbar^2} \left(A - B\sqrt{1 + \frac{C^2}{3B^2}} \right). \end{aligned} \quad (2.173)$$

- **Direction [100].** One has

$$\begin{cases} \frac{1}{m_+} = \frac{2}{\hbar^2} (A + B) \\ \frac{1}{m_-} = \frac{2}{\hbar^2} (A - B) \end{cases}. \quad (2.174)$$

- **Valence band isotropic mass.** This is obtained after averaging over all possible \mathbf{k} direction. The result is:

$$\begin{aligned} \frac{1}{m_+} &= \frac{2}{\hbar^2} \left[A + \frac{2}{5}B \left(1 + \frac{3}{2}\sqrt{1 + \frac{4C^2}{9B^2}} \right) \right], \\ \frac{1}{m_-} &= \frac{2}{\hbar^2} \left[A - \frac{2}{5}B \left(1 + \frac{3}{2}\sqrt{1 + \frac{4C^2}{9B^2}} \right) \right], \end{aligned} \quad (2.175)$$

where the parameter values are: $A = -4.28$, $B = -0.68$, and $C^2 = 24$ in $\hbar^2/(2m_e)$ units. One immediately sees that $m_+ = -0.15$ and $m_- = -0.50$.

Gallium Arsenide. The band structure calculation for GaAs requires the determination of both antisymmetric and symmetric form factors. From a fit of spectroscopic data one obtains the data shown in [Table 2.2](#).

The band structure generated after using these parameters is displayed in [Fig. 2.34](#). A few differences are to be remarked in comparison with Si band structure in [Fig. 2.33](#):

1. The gap is direct and occurs at the BZ center

Table 2.2 Band structure of GaAs from semi-empirical pseudopotential method. Values of V_S and V_A (Ry) that fit the experimental data

| $V_S(\mathbf{G}_3)$ | $V_S(\mathbf{G}_8)$ | $V_S(\mathbf{G}_{11})$ | $V_A(\mathbf{G}_3)$ | $V_A(\mathbf{G}_4)$ | $V_A(\mathbf{G}_{11})$ |
|-----------------------|-----------------------|--------------------------|-----------------------|-----------------------|--------------------------|
| -0.2396 | 0.0126 | 0.06 | 0.07 | 0.05 | 0.01 |

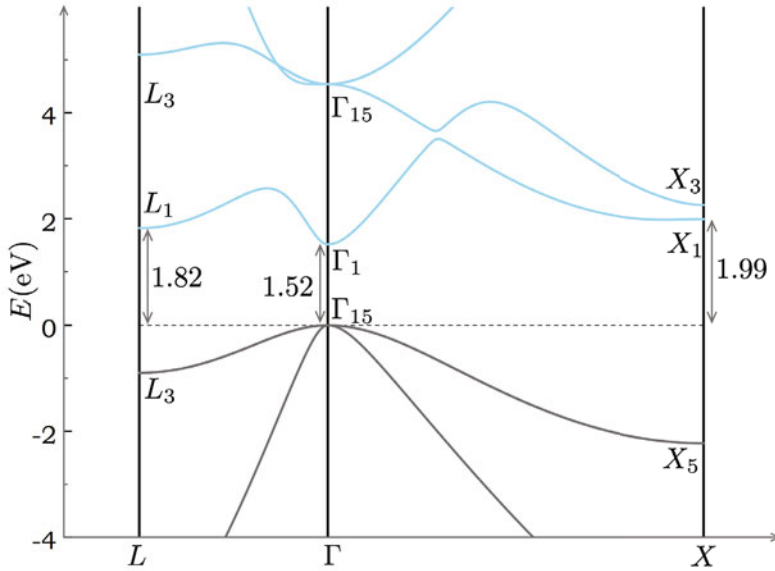


Fig. 2.34 Band structure of GaAs from semi-empirical pseudopotential method. Labels for the symmetry points and corresponding directions in the first BZ are those depicted in Fig. 1.29

2. The conduction band minimum is not degenerate. In the vicinity of $\mathbf{k} = \mathbf{0}$ one can thus expand $E(\mathbf{k})$ and obtain

$$E(\mathbf{k}) = E_g + \frac{\hbar^2}{2m_c} (k_x^2 + k_y^2 + k_z^2), \quad (2.176)$$

with $m_c = 0.067m_e$ and $E_g = 1.5$ eV.

3. While considerations similar to the Si case hold, the parameter values here are instead $A = -6.9$, $B = -4.4$, and $C^2 = 43$ in units of $\hbar^2/(2m_e)$. As a result, one has $m_+ = -0.08$ and $m_- = -0.70$

2.5.6 Excitonic effects

In the model represented by (2.156) and discussed in Sec. 2.5.3, the presence of electrostatic and exchange energy terms among valence electrons have been found to lead to an interesting consequence: in the presence of many electrons, the operation of promoting one electron from a state at lower energy ε_1 to a state with higher energy ε_2 , requires an amount of energy smaller than energy separation between the two levels, that is $\varepsilon_2 - \varepsilon_1$. This effect has been traced back to the interplay between repulsive electrostatic and attractive exchange energies. The effect

has emerged while analyzing systems with a few valence electrons or with a few holes, namely with valence states close to full occupation. This is known as excitonic effect and has crucial consequences on the optical and transport properties of crystals and nanostructures. A complete treatment of this problem is quite complicated within a many-particle theory framework. In spite of its complexity, its origin can be intuitively understood.

Consider to have a semiconductor at zero temperature. As already discussed, this corresponds to a situation in which all the energy bands below the valence band states are fully occupied along with the valence band itself. Within the Hartree-Fock method this means that the self-consistent single-particle field has been constructed, yielding all the electronic states of the N valence electrons. The conduction band states are then calculated after solving a single-particle Schrödinger equation where the potential energy is self-consistently determined for the fundamental, i.e. lowest-energy, state. Actually, a problem with $N + 1$ electrons is solved.

Consider now the case in which an extra electron is injected in the semiconductor. The situation illustrated above is then realized, and the new electron occupies the lowest state in the conduction band. A different situation might occur instead if, for example, an electron is promoted from the valence to the conduction band. In this case excitonic effects arise, precisely in the way explained so far. Promotion of the electron from valence to conduction band clearly modifies the self-consistent field. In order to restore it, the self-consistent field of the promoted electron is to be added. Not enough, it has to be finally subtracted in order to leave the total potential unchanged. This last term is equivalent to the potential contribution from a positive charge. The complete theory leads to state that the part of wavefunction describing the correlation between electron and hole is described by the Hamiltonian

$$H = E_g - \frac{\hbar^2}{2m_e^*} \nabla_{\mathbf{r}_e}^2 - \frac{\hbar^2}{2m_h^*} \nabla_{\mathbf{r}_h}^2 - \frac{e^2}{\epsilon_0 |\mathbf{r}_e - \mathbf{r}_h|}, \quad (2.177)$$

where E_g is the energy gap separating the maximum of valence band from the minimum of conduction band. Besides, m_e^* and m_h^* are respectively electron and hole effective masses as determined in the vicinity of, again respectively, the minimum of conduction band and the maximum of valence band. Finally, ϵ_0 is an appropriate dielectric constant accounting for the polarizability of the medium. Equation (2.177) holds within the effective mass approximation, that is whether the wavefunction is delocalized over several primitive cells. It describes the formation of an electron-hole pair as a result of a Coulomb potential, the pair center-of-mass executing a free motion. The excited states of this N -electron system described by the Hamiltonian (2.177) is called Wannier exciton.

Once the center-of-mass motion with momentum $\hbar \mathbf{K}$ is separated, the energies of the states are

$$E_n(K) = E_g + \frac{\hbar^2}{2M} K^2 - \frac{Ry}{n^2}. \quad (2.178)$$

The energy spectrum is composed of a free-particle term $\hbar^2 K^2 / (2M)$ describing the center-of-mass motion, with $M = m_e^* + m_h^*$ the total effective mass of electron and

hole, and a binding energy term having the same form as that of an hydrogen atom $-Ry/n^2$, with n the principal quantum number assuming all integer values equal or larger than 1. The difference with the hydrogen atom spectrum is expressed in effective Rydberg, that is $Ry = e^2/(2a_b^{\text{eff}})$: this is defined in terms of the effective Bohr radius $a_b^{\text{eff}} = \epsilon_0 \hbar^2 / (\mu e^2) = \epsilon_0 m a_b / \mu$ containing the hydrogen Bohr radius a_b , the free electron mass m and the effective electron-hole reduced mass $\mu^{-1} = (m_e^*)^{-1} + (m_h^*)^{-1}$. The effective Bohr radius is about 100 times larger than Bohr radius, due to the fact that parameters values typical in semiconductors such Si and GasAs, are $\mu \sim 0.1m$ and $\epsilon_0 \sim 10$. Thus, the binding energy of an exciton as referred to the conduction band minimum can be as small as a few tens of meV while a_b^{eff} is of the order of 50 \AA . Excitonic effects in a solid thus determine the presence of bound electron-hole pairs, whose wavefunction extends over tens of lattice sites and binding energies correspondingly smaller than those of hydrogen atom. This finding justifies a posteriori the use of effective masses.

The energy of an excitonic state can be measured in experiments of high-frequency light absorption, as is discussed in detail in Chap. 6. For illustration purposes, Fig. 2.35 displays the absorption spectrum for a ZnSe crystal, that is the intensity of absorbed light as a function of the energy of incoming photons. Part (a) displays the absorption spectrum at low temperatures T , namely such that $k_b T \ll Ry$: a well defined absorption spectrum appears below the energy gap E_g . Part (b) displays the data for higher temperatures such that $k_b T \gg Ry$: the spectrum below E_g is flattened due to the fact that excited exciton states with $n > 1$ are involved.

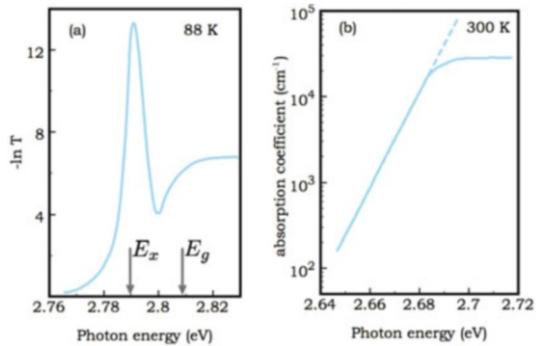


Fig. 2.35 Excitonic effects. Absorption spectrum of ZnSe. For comparison, data displayed for (a) low temperatures $T \ll Ry/k_b$, and (b) high temperature $T \gg Ry/k_b$ [17]

Quick Questions

- Q12.** The description of excitonic effects within the Wannier framework requires that electron and hole be at large average distance on the lattice parameter scale. Discuss whether in a crystal with only one atom per cell, a kind of excitonic effect be imagined in which the electron-hole pair is localized at the atom where the electron has been excited from.

Answer. There is no reason to think that the exciton effects is not present when the electron and the hole are separated at distance of the order of the atomic radius. These are called Frenkel excitons

Q13. If the center-of-mass motion is taken into account, does the exciton energy spectrum become discrete?

Answer. No. For each given bound excitonic state, the spectrum has a parabolic continuum form with respect to the center-of-mass wavevector K .

2.6 Band structures in quantum confined nanostructures

Quantum confinement of charge carriers, that can be holes and electrons, can be realized in one, two, and three dimensions. The shape and size can be engineered to drive wanted quantum-mechanical effects and to precisely control transport properties. Quantum confinement of particles can be theoretically described at least by two complementary approaches as schematized in Fig. 2.36. In atomistic approach, the microscopic structure of underlying ion cores is taken into account by explicitly including in the Hamiltonian the electron-ion interaction potential. In effective-mass approach instead, electron (hole)-ion interactions are embodied by assigning to the particle, electron or hole, a band index and an effective mass, according to the concepts developed in Sec. 2.3.3. The latter embodies the effect of the underlying ion cores, though in an effective and approximated way. In fact, the atomistic approach could be employed in much the same way as already done while developing the theory of electronic bands for one-dimensional systems such as atomic chains and carbon nanotubes. However, the effective-mass approach is considered in the following, because it allows to catch the main concepts while using analytical derivations. The same approach is used in the next section to investigate exciton formation in nanometric systems.

2.6.1 Single-particle states

The electronic spectrum and wavefunctions in nanostructures can be determined according to the following

Procedure

Step 1. Identify the confinement potential $V_{\text{conf}}(\mathbf{r})$ that properly describes the confinement of the particle. A common choice is the hard-wall potential

$$V_{\text{conf}}(\mathbf{r}) = \begin{cases} 0 & \text{inside the nanostructure} \\ +\infty & \text{otherwise} \end{cases} \quad (2.179)$$

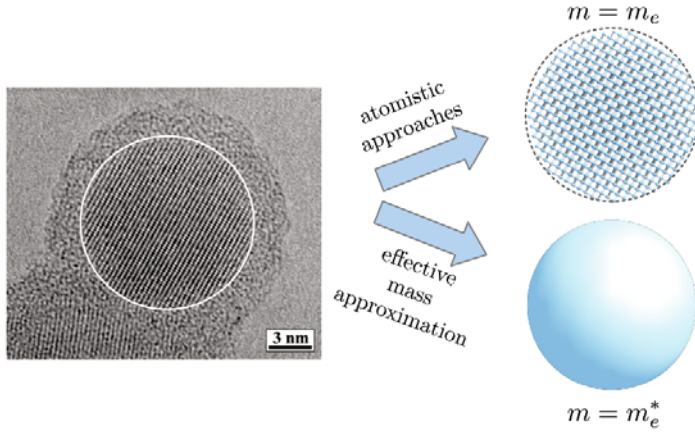


Fig. 2.36 Quantum confinement in nanostructures. Two possible theoretical approaches are schematized: atomistic and effective-mass approaches (see text). The image on the left is from experiment [18]

This choice corresponds to a zero probability of finding the particle at the nanostructure boundaries. Notice that other choices are of course possible, like assuming a finite potential barrier at the boundary.

Step 2. Identify the boundary condition for the electron or hole wavefunction $\Psi(\mathbf{r})$, that be compatible with the confinement potential. In the case of a hard-wall potential, this is

$$\Psi(\mathbf{r})|_{\mathbf{r} \in \text{boundary}} = 0.$$

Step 3. Solve the Schrödinger equation

$$\frac{-\hbar^2}{2m^*} \nabla^2 \Psi(\mathbf{r}) + V_{\text{conf}}(\mathbf{r}) \Psi(\mathbf{r}) = E \Psi(\mathbf{r}),$$

with the boundary condition fixed at Step 2.

Examples

Nanostructures can be classified according to their dimensionality. In the following, the above Procedure is applied to illustrate the electronic band structure of quantum wells (2D), quantum wires (1D) and quantum dots (0D) nanostructures within the simple effective-mass approximation.

Quantum wells. These are formed by embedding a layer of lower band-gap materials between two layers of a higher band-gap material. The band gap mismatch works as a potential well for charge carriers, that is as a confinement potential. The quantum confinement is thus active in one direction, leading to two-dimensional nanostructures, as depicted in Fig. 2.37. A useful parameter here is the band offset ratio, that is the ratio of the conduction to the valence bands energies. Two situations may occur: either the conduction (valence) band edge in one material is lower (higher) than the corresponding edge in the other, or both the conduction and valence band edges in one material are lower than the corresponding band edges in the other material. The former are called type I heterostructures, where electrons and holes can both live in one material, and the latter type II heterostructures, where electrons and holes are separately confined in the two materials. Type I heterostructures include III-V compounds such as GaAs/GaAlAs, GaSb/GaAlSb, InGaAs/InAlAs. Type II heterostructures include InAs/AlGaSb, InAs/GaSb. The effective mass approximation leads to the solution of the Schrödinger equation in a infinite planar slab with thickness L along the z -direction. Application of the Procedure outlined above results into the eigenfunctions and eigenvalues:

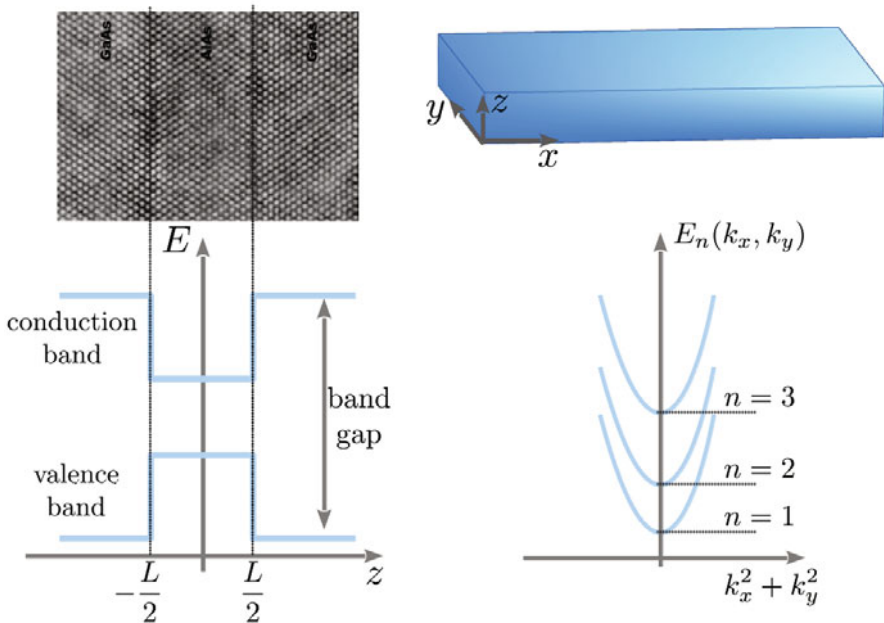


Fig. 2.37 Left panel: a typical quantum well structure [19], along with the corresponding conduction and valence band-edge profiles. The z direction is orthogonal to the two interfaces and represents the confinement direction. The charge carriers are free in x and y directions. Right panel: the miniband structure is represented (bottom). This is composed of parabolic bands corresponding to free motion with effective mass along the slab's x and y directions (top). The parabolic bands are edging on discrete energy levels, that correspond to confinement in the z direction. The miniband edge positions are the signature of quantum confinement

$$\Psi_n(x, y, z) = A \sqrt{\frac{2}{L}} e^{ik_x x} e^{ik_y y} \sin\left(\frac{n\pi}{L} z\right)$$

$$E_n(k_x, k_y) = \frac{\hbar^2}{2m^*} \frac{n^2 \pi^2}{L^2} + \frac{\hbar^2}{2m^*} (k_x^2 + k_y^2), \quad (2.180)$$

with A a normalization constant. Therefore, carriers move within the slab x and y directions as free particles, with effective masses and thus according to parabolic band dispersion. Confinement along the z direction originates the discrete level structure, whose edges are located at $\hbar^2/(2m^*)(n^2 \pi^2/L^2)$, corresponding to the discrete eigenvalues. This structure is named miniband. The present model evidently assumes that the effective mass be isotropic.

Quantum wires. These are one-dimensional nanostructures, where the quantum confinement is active in two different directions, the third one remaining free. The electronic bands depend on the quantum confinement effects via the wire shape and especially section, that can be modeled as squared, rectangular, circular, and so on. Application of the above Procedure to a long circular cylindrical sample with axis along the z -direction, yields the eigenfunctions and eigenvalues:

$$\Psi_n(\rho, z) = A \frac{1}{\sqrt{2\pi}} e^{ik_z z} e^{im\varphi} J_m\left[\frac{\chi_{m,p}}{R} \rho\right]$$

$$E_{m,p}(k_z) = \frac{\hbar^2}{2m^*} k_z^2 + \frac{\hbar^2}{2m^*} \frac{\chi_{m,p}^2}{R^2}, \quad (2.181)$$

with A again a normalization constant. In addition, R is the nanowire radius, J_m is the m -th cylindrical Bessel function, $\chi_{m,p}$ its p -th zero, ρ the radial coordinate in the $x-y$ plane. The resulting electronic spectrum is qualitatively similar to that of a quantum well: a miniband structure shows up indeed, where now the miniband edges depend on the nanowire size and section shape. Also, the minibands are one dimensional in character, since carriers are free only along the z direction.

Quantum dots or nanocrystals. Quantum confinement in all directions results into a zero-dimensional nanostructure or quantum dot. The simplest situation occurs when the quantum dot can be modeled as spherical with radius R . In this case the wavefunction can be expressed in terms of the product of a radial function, that is a spherical Bessel function, times an angular function, that is given by spherical harmonics. In other words, application of the Procedure yields:

$$\Psi_{n,l,m} = A j_n\left(\frac{\chi_{nl}}{R} r\right) Y_{l,m}(\theta, \varphi),$$

where $n = 0, 1, 2, \dots$, $l = 0, 1, 2, \dots$, and $m = -l, \dots, 0, \dots, l$ label the principal, orbital angular, and azimuthal angular momentum quantum numbers, respectively. The corresponding spectrum is

$$E_{n,l} = \frac{\hbar^2 \chi_{n,l}^2}{2m^* R^2}, \quad (2.182)$$

with $\chi_{n,l}$ the l -th zero of the spherical Bessel function of order n . $E_{n,l}$ are confinement energies: they are originated from constraining the wavefunction inside the dot, so that they result to be positive. Along these lines, the smaller is the dot size, the larger are the confinement energies according to a R^{-2} scaling law. A schematic representation of the lowest-energy eigenvalues (2.182) is given in Fig. 2.38. Notice that the discrete electronic spectrum of these nanostructures is quite similar to that of atoms and molecules. For this reason quantum dots are also called artificial atoms.

In real labs, nanodots with different shapes can be engineered, so that the possibility of determining the confinement energy may become uneasy. For example, in the case of ellipsoidal dots such as those realized in [20], the measured confinement energy results to be larger than that of a corresponding spherical dot with equal volume. Besides, the wavefunction is found to be localized along special directions. Both expressions (2.182) for a 3D spherical dot and (2.45) for a 1D linear chain show that level separation and band gaps values increase with decreasing the system size. It can be demonstrated that anisotropies in masses and confinement geometry positively contribute to the confinement energy, as represented in Fig. 2.39. This bears important consequences in optical and transport properties [21].

A few final remarks are in order when assessing the validity of the approximations adopted to determine electronic spectra in quantum dot structures. First, if λ_e and λ_h are the electron and hole de Broglie wavelengths respectively, the effective mass approximation fails in dots with small sizes $R \ll \lambda_e$ or $R \ll \lambda_h$, so that calculations from first principles are needed to determine the energy levels. Second, the crystal might not be amenable to a single energy-band description. As an example, the valence band of silicon is fourfold degenerate in proximity of its maximum, so that the effective masses are anisotropic. Last point, the modeling of confinement potential by an infinite barrier might be too crude to be realistic.

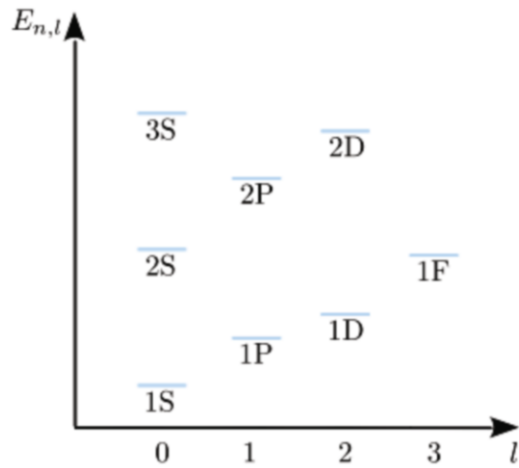
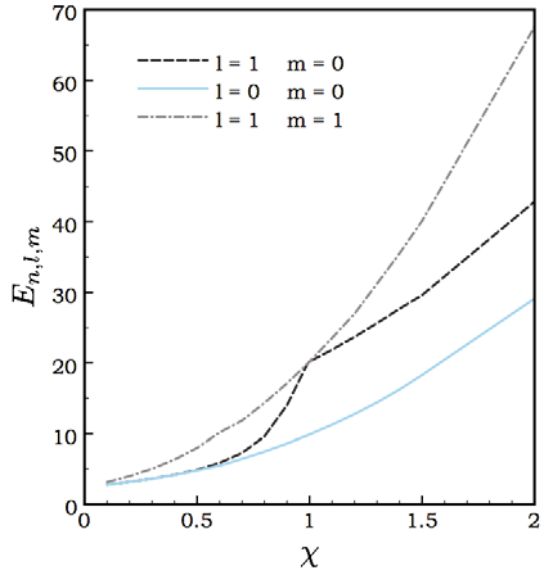


Fig. 2.38 Quantum confinement in spherical quantum dots. A schematic representation of the eigenvalues (2.182). The notation is similar to that labeling the energy levels in atoms. The integer numbers label the principal quantum number and the progressive letters S, P, D, and F label the value of the orbital angular $l = 0, 1, 2, 3$ quantum number, respectively. Capital letters are conventionally used

Fig. 2.39 Quantum confinement in ellipsoidal quantum dots. Electronic energies as a function of the aspect ratio $\chi = c/a$ for three different states: ground state with $n = l = m = 0$ (solid line), first two excited states with $l = 1$ and $n = m = 0$ (dashed line) and with $n = 0, l = |m| = 1$ (dot-dashed line). $a = b$ and c are the ellipsoid semiaxes and the energies are in units $\hbar^2/2m^*c^2$ [21]



2.6.2 Excitons in nanometric systems

The study of excitonic effects in confined structures needs a precise assessment of the different length scales governing the system. These are the lattice parameter a , the de Broglie wavelength of either electron or hole, say λ_e and λ_h respectively, the effective exciton Bohr radius a_b^{eff} , and the typical linear size R of the given dot. For the effective mass approximation to remain valid, the conditions $\lambda_e \gg a$ and $\lambda_h \gg a$ are to be fulfilled [2].

Consider a quantum dot as in the previous section, with spherical geometry and isotropic effective mass. Two opposite regimes can be met. The exciton can be weakly confined, a situation that can be expressed in mathematical form as $\lambda_e \gg a$ and $\lambda_h \gg a$ and $a_b^{\text{eff}} \ll R$. Within the effective mass approximation, $\lambda_e \ll R$ and $\lambda_h \ll R$, and the exciton can be tightly confined if $a_b^{\text{eff}} \gg R$. The energy spectra have different characteristics in the two regimes, as it is being explored in the following.

Weakly confined exciton. In this regime the solution of the Schrödinger equation associated to Hamiltonian (2.177) is obtained considering that: a) since $a_b^{\text{eff}} \ll R$, the exciton problem can be solved disregarding the confinement; b) the center of mass motion already embodies the confinement. As a consequence, the energy spectrum is obtained:

$$E_{nqp} = E_g - \frac{Ry}{n^2} + \frac{\hbar^2 \chi_{q,p}^2}{2MR^2}, \quad (2.183)$$

with $\chi_{q,p}$ being the q -th zero of the spherical Bessel function of order p . The hydrogenoid electron-hole state with energy $-Ry/n^2$ is labeled by three quantum numbers: the principal n , the orbital angular momentum l of the particle with reduced mass μ , and the azimuthal m . The confinement energy of the center of mass is $\hbar^2 \chi_{q,p}^2 / (2MR^2)$, and is characterized by angular momentum p , an azimuthal quantum number, and an integer q . The following spectroscopic-like notation is commonly used: a) for electron-hole states the symbols ($1S$, $2S$, $2P$, $3S$, $3P$, $3D$, ...); b) for the confinement term ($1s$, $2s$, $2p$, $3s$, $3p$, ...).

The fundamental state $E_{1S,1s}$ with $n = 1$, $q = 1$, and $p = 0$ has energy

$$E_{1S,1s} = E_g - Ry + \frac{\hbar^2 \pi^2}{2MR^2} = E_g - Ry \left[1 - \frac{\mu}{M} \left(\frac{\pi a_b^{\text{eff}}}{R} \right)^2 \right]. \quad (2.184)$$

It is therefore seen that the fundamental level acquires a positive shift with respect to the case of an exciton in an infinite crystal, that is given by

$$\Delta E_{1S,1s} = Ry \frac{\mu}{M} \left(\frac{\pi a_b^{\text{eff}}}{R} \right)^2, \quad (2.185)$$

This is a tiny shift on the scale of Ry , since $\mu < M$ and $a_b^{\text{eff}} \ll R$.

If Coulomb interactions between electron and hole were negligible, the single-particle energies are

$$E_{q,p}^e = E_g + \frac{\hbar^2 \chi_{q,p}^2}{2m_e^* R^2}, \quad \text{and} \quad E_{q,p}^h = \frac{\hbar^2 \chi_{q,p}^2}{2m_h^* R^2}. \quad (2.186)$$

The total energy of the lowest, fundamental state is ($q = 1$, $p = 0$)

$$E_g^{\text{eff}} = E_g + \frac{\hbar^2 \pi^2}{2\mu R^2} = E_g + Ry \left(\frac{\pi a_b^{\text{eff}}}{R} \right)^2 = E_g^{\text{eff}}. \quad (2.187)$$

The effective energy gap E_g^{eff} that has just been introduced embodies the localization of the single-particle states. With this definition, the binding energy can be expressed in terms of

$$Ry^{\text{eff}} = Ry \left[1 + \left(1 - \frac{\mu}{M} \right) \left(\frac{\pi a_b^{\text{eff}}}{R} \right)^2 \right] > Ry. \quad (2.188)$$

To conclude, the electronic hydrogenoid level of a weakly confined exciton is shifted towards larger energies than in the infinite crystal. As a result, the weakly bound exciton binding energy is smaller than for an exciton in infinite crystal.

Strongly confined exciton. In this case with $R \ll a_b^{\text{eff}}$, exciton formation is prevented. Thus (2.183) is no longer valid. One has to revert back to the set (2.186) which, for the same values of q and p , provide the kinetic energies of the electron-hole pair. An easy summation of the two equations in (2.186) yields:

$$E_{q,p} = E_g + \frac{\hbar^2 \chi_{q,p}^2}{2\mu R^2}. \quad (2.189)$$

Under this assumption, the electron and the hole behave as two independent particles. Indeed, it turns out that under strong confinement, the kinetic energy (2.189) represents the dominant contribution to the total energy. However, the Coulomb attraction can be treated as a perturbation to the independent particles ground state. One finds [22]

$$E_{1s,1s} = E_g + Ry \left(\frac{\pi a_b^{\text{eff}}}{R} \right)^2 - 1.786 Ry \frac{a_b^{\text{eff}}}{R} - 0.248 Ry. \quad (2.190)$$

The second term on the right-hand-side represents the confinement energy of the electron-hole pair without electrostatic interaction. The third represents the electrostatic contribution, and the fourth, independent of the dimension of the dot, reduces the energy of the pair with respect to the result (2.189) and represents a correlation effect. Equation (2.190) clearly shows that, on reducing the dimension, the kinetic energy contribution becomes dominant, with a scaling law $\propto R^{-2}$, to be contrasted with the Coulomb contribution, which scales as R^{-1} . Fig. 2.40 displays the energy $(E_{1s,1s} - E_g)/Ry$ vs. a_b^{eff}/R , as calculated in [23]. Different materials correspond to different parameters and thus different values of a_b^{eff}/R . Thus, the conclusion can be drawn that the curve is very little dependent on the given material, at least beyond a threshold.

A complete study of Hamiltonian (2.177) leads to the conclusion that the analytical spectra (2.184) and (2.190) are generally in good agreement with the results of more refined calculations whenever $R/a_b^{\text{eff}} > 4$ and $R/a_b^{\text{eff}} < 1$.

Excitons in ellipsoidal quantum dots have also been investigated, but the complexity of the corresponding calculations puts them beyond the scope of this textbook. The interested student can be addressed to several references, an useful starting point being [24].

Quick Questions

- Q14.** Provide a qualitative description of the evolution of an excitonic system while the crystal size decreases down to nanometric sizes, in particular in terms of the evolution of its different energy components and of its size.

Answer. Excitonic spectra in an infinite crystal contain a hydrogen-like binding energy summed to a center-of-mass kinetic energy term of the electron-hole pair. A decrease of crystal size first leads to a tiny decrease of the pair binding energy. This decreasing behavior keeps going until the dot size is comparable to the effective size a_b^{eff} of the exciton. At this point, the correlated electron-hole system breaks up and the confinement energy term becomes dominant for the two particles. In turn, the electrostatic energy in turn increases while the dot gets smaller.

2.7 Spin-related magnetic ground states and excitations

2.7.1 Introduction and basic facts

Materials are composed of charged particles which possess a spin and move around. Therefore, materials possess intrinsic magnetic properties, which can be enhanced or modified in the presence of external magnetic fields. Here a schematic illustration of the main concepts is given, while more extended presentations can be found in [4, 25]. In order to be self-contained, basic knowledge on magnetism is first schematically recalled from elementary physics.

In essence,

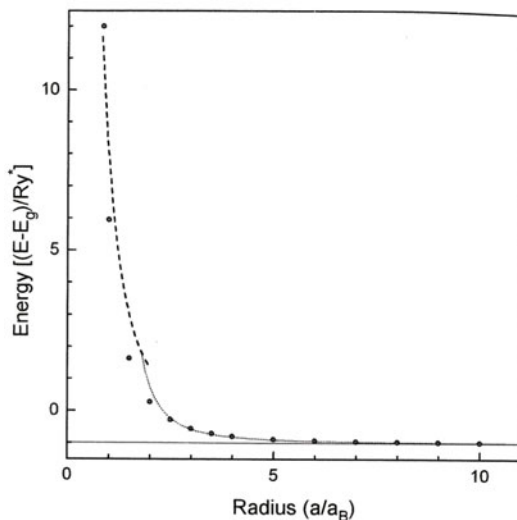


Fig. 2.40 Exciton formation in strongly confined quantum dots. Calculated $(E_{1s,1s} - E_g)/Ry$ as a function of a_b^{eff}/R [23]. Note that beyond a threshold, the curve is very little dependent on the given material. Reprinted with permission from [23]. Copyright (1988) by the American Physical Society

Concept

Magnetism is a phenomenon due to moving charges. Electrons in atoms or molecules composing the material produce small magnetic fields as they spin and move around the atom. Thus, magnetic effects arise from interactions between these moving charges or with external magnetic fields.

The angular momentum and spin of nuclei is also to be considered, though the heavier mass of nuclei appearing in the Bohr magneton suggests that their magnitude can be neglected to a first approximation.

Since the moving charges in atoms are sources of magnetic fields, the question is how materials respond to the presence of an applied magnetic field. This response is measured by magnetic susceptibilities χ_m : after the application of a magnetic field \mathbf{H} , the material builds up a magnetization $\mathbf{M} = \chi_m \mathbf{H}$ that is the magnetic dipole moment per unit volume. On the other hand, the induced magnetic field $\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M}$ in the material results from the external and the magnetization parts, so that $\mathbf{B} = (1 + 4\pi\chi_m)\mathbf{H}$. Materials are considered diamagnetic if the susceptibility is negative and paramagnetic if it is positive. In both cases magnetic susceptibilities are usually very small $\ll 1$. Materials exist that have very large susceptibilities: these are for example ferromagnets and antiferromagnets.

In diamagnetic materials, the atomic magnetic moment cancel out when the field is off, but a magnetic moment is induced after the application of the magnetic field, which the induced magnetic moment opposes against. In paramagnetic materials instead, a finite intrinsic magnetic moment is present from the beginning: this is then oriented by the applied magnetic field so to produce an additional magnetization. The latter depends on the statistical distribution of the magnetic moment orientations, and thus on temperature. Considered the classification above, all insulators turn out to be diamagnetic while metals and semiconductors show up both paramagnetic and diamagnetic behavior.

While diamagnetism and paramagnetism are essentially phenomena originated by interactions of single magnetic dipoles of atoms with a magnetic field, quite remarkable new ground states can be originated by the presence of correlations among atomic magnetic dipoles and lead to macroscopic collective phenomena. Ferromagnetic and antiferromagnetic materials are two such examples, in which below a given critical temperature a new ground state is energetically favorite. Under these conditions, a macroscopic alignment of magnetic moments permanently manifests even in the absence of the magnetic field: with parallel alignment in ferromagnets so to have a finite macroscopic permanent magnetic moment, antiparallel with different sizes and nonzero net magnetic moment in ferrimagnets, and antiparallel with equal sizes and zero net magnetic moment in antiferromagnets.

In the following, a formal development is performed on these matters, to provide quantitative investigation tools.

2.7.1.1 Elementary magnetic dipoles

Consider an elementary electric current i on a close loop delimiting a surface $\mathbf{S} = S\hat{\mathbf{n}}$, that is oriented perpendicular to direction $\hat{\mathbf{n}}$. From elementary physics it is known that the interaction of the current i with a magnetic field \mathbf{B} is equivalent to that of a magnetic dipole $\boldsymbol{\mu} = iS\hat{\mathbf{n}}/c$ with \mathbf{B} , with c the speed of light. In essence, a loop current is equivalent to a magnetic dipole.

Application of this concept to atoms, yet within a classical framework, leads to the idea that electrons moving around the nucleus with angular frequency ω_0 produce a current $i = -e\omega_0/(2\pi)$: therefore, such an elementary atomic electron current can be represented as a magnetic dipole:

$$\boldsymbol{\mu} = -\frac{e\omega_0}{2\pi} \frac{\pi r^2}{c} \hat{\mathbf{n}} = -\frac{e\omega_0 r^2}{2c} \hat{\mathbf{n}}. \quad (2.191)$$

Here, r represents the radius of the circular orbit centered on the nucleus. Since the particle's angular momentum is $\boldsymbol{\ell} = \mathbf{r} \wedge \mathbf{p} = m\omega_0 r^2 \hat{\mathbf{n}}$, one has

$$\boldsymbol{\mu} = -\frac{e}{2mc} \boldsymbol{\ell}, \quad (2.192)$$

valid even for non-circular trajectories. Expression (2.192) represents the connection between orbital angular momentum ℓ and magnetic dipole concepts, at least within a classical framework.

In fact, it can be demonstrated that Eq. (2.192) holds as well in quantum world, provided that $\boldsymbol{\ell}$ be replaced by the orbital angular momentum operator $\hat{\boldsymbol{\ell}}$. One thus obtains:

$$\begin{aligned} |\boldsymbol{\mu}| &= \sqrt{\ell(\ell+1)} \frac{e\hbar}{2mc}, \\ \mu_z &= m_\ell \frac{e\hbar}{2mc}, \end{aligned} \quad (2.193)$$

with ℓ and m_ℓ integers such that $-\ell \leq m_\ell \leq \ell$. The quantity $\mu_b = e\hbar/(2mc)$ is called Bohr magneton.

Besides orbital angular momentum, particles do have a spin as well. Following the procedures known from quantum mechanics to embody the spin operator too, the operator $\hat{\mu}_z$ becomes:

$$\hat{\mu}_z = \frac{\mu_b}{\hbar} (\hat{\ell}_z + 2\hat{s}_z), \quad (2.194)$$

with $\hat{\ell}_z$ and \hat{s}_z the operators for the z -components of orbital angular and spin momenta, respectively. The factor of 2 results from quantum theory and is experimentally verified.

Passing from one to the many electrons in atoms requires the introduction of the total orbital \mathbf{L} and spin \mathbf{S} angular momenta. Two cases are in order. If \mathbf{L} and \mathbf{S} are separately good quantum numbers, then the total magnetic momentum is

$$\mu_z = \mu_b(M_L + 2M_S). \quad (2.195)$$

If spin-orbit interaction is taken into account, then the good quantum numbers are related to the total angular momentum operator $\mathbf{J} = \mathbf{L} + \mathbf{S}$, and to the operators \mathbf{J}^2 , \mathbf{L}^2 , \mathbf{S}^2 . The corresponding eigenvalues are $J(J+1)\hbar^2$, $L(L+1)\hbar^2$, and $S(S+1)\hbar^2$, with $-J \leq M_J \leq J$ and the M_J values being $2J+1$ numbers integer or half-integer. Therefore, one has that

$$\mu_z = \mu_b g M_J, \quad (2.196)$$

in terms of the so-called Landé factor

$$g = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)}. \quad (2.197)$$

2.7.1.2 Density of magnetic energy

It is known from elementary physics that the energy of a magnetic dipole $\boldsymbol{\mu}$ in a field \mathbf{B} is $-\boldsymbol{\mu} \cdot \mathbf{B}$. Thus, to first order in χ_m , a system in a magnetic field acquires the density of magnetic energy:

$$U = - \int_0^B \mathbf{M} \cdot d\mathbf{B} = -\frac{1}{2} \chi_m B^2. \quad (2.198)$$

When the magnetic field is not uniform, the force acting in the z -direction amounts to:

$$F_z = - \frac{\partial U}{\partial z} = \chi_m B \frac{\partial B}{\partial z}. \quad (2.199)$$

A measure of the force yields a way to experimentally determine χ_m .

2.7.2 Diamagnetism

Expressions (2.195) and (2.196) represent the magnetic dipoles originated by orbital and spin angular momenta. When placed in an external magnetic field \mathbf{B} along, say, the z direction $\mathbf{B} = B\hat{\mathbf{z}}$, the resulting interaction energy is $U = -\mu_z B$. Consider atoms with completely filled shells. They have $L = S = 0$ and thus $J = 0$: in this case the magnetic dipole is zero and no interaction takes place with the external magnetic field. Manifestations of such a condition are for example molecular solids and ionic crystals, like rare-gases and alkali halides crystals, respectively.

However, at least within a classical-mechanics framework, the magnetic field modifies the electron's angular velocity by the Larmor term $\omega_L = -eB/(2mc)$. Considering that the atom has Z electrons, (2.191) therefore leads to a new expression for the magnetic-dipole:

$$\mu_L = \omega_L \frac{Ze}{2c} \langle \rho^2 \rangle, \quad (2.200)$$

where $\langle \rho^2 \rangle$ is the average squared radius of electron trajectories in a plane perpendicular to \mathbf{B} . If N is the density of atoms, the magnetization turns out to be

$$M = \mu_L N = -N \frac{Ze^2}{4mc^2} \langle \rho^2 \rangle B. \quad (2.201)$$

The magnetic susceptibility is thus

$$\chi_m = \frac{M}{B} = -\frac{Ze^2}{6mc^2} \langle r^2 \rangle N, \quad (2.202)$$

where for simplicity the case of isotropic systems has been considered, so that $\langle \rho^2 \rangle = \langle x^2 + y^2 \rangle = 2\langle r^2 \rangle / 3$ and the factor 6 comes about.

In essence,

Concept

In systems with zero total angular momentum the magnetic moments cancel out in the absence of a magnetic field. However, when the field is switched on, the system responds opposing against the field via a magnetic susceptibility $\chi_m < 0$, whose magnitude $|\chi_m|$ increases with increasing $\langle r^2 \rangle$.

This is indeed experimentally found in atoms or ions with completely filled shells, that are characterized by very small values of $|\chi_m| \approx 10^{-6}$, the actual value increasing with the atom size. This phenomenon is called diamagnetism.

If the same calculations were performed within a quantum-mechanics framework, not much would be changed. In quantum mechanics, the Hamiltonian of a single particle in a magnetic field and the potential $V(\mathbf{r})$ is

$$H = \frac{1}{2m} \left(\mathbf{p} - \frac{e}{c} \mathbf{A} \right)^2 + V(\mathbf{r}), \quad (2.203)$$

with \mathbf{A} the vector potential related to the magnetic field \mathbf{B} . If $\mathbf{B} = B\hat{\mathbf{z}}$, \mathbf{A} can be cast in the form $\mathbf{A} = \mathbf{B} \wedge \mathbf{r} / 2 \equiv (-By, Bx, 0) / 2$, and (2.203) becomes

$$H = \frac{p^2}{2m} + V(\mathbf{r}) - \frac{eB}{2mc} l_z + \frac{e^2 B^2}{8mc^2} (x^2 + y^2). \quad (2.204)$$

Treating the last term on the right-hand side to first order in perturbation theory, one obtains

$$\Delta E = \frac{e^2 B^2}{8mc^2} \langle \psi | x^2 + y^2 | \psi \rangle, \quad (2.205)$$

where ψ is the unperturbed electron wavefunction. Considering N atoms per unit volume and identifying (2.205) with (2.198), one has

$$\chi_m = -\frac{2\Delta E}{B^2} = -\frac{e^2 N}{6mc^2} \sum_i \langle \psi_i | r^2 | \psi_i \rangle. \quad (2.206)$$

Here, the sum is executed over all the electronic wavefunctions of the i -th atom, and $\langle \psi | x^2 + y^2 | \psi \rangle = 2\langle r^2 \rangle / 3$ represents the quantum-mechanical average of the squared position in an isotropic medium. In fact, this is the only difference between the classical (2.202) and the quantum (2.206) expressions for χ_m , otherwise formally equivalent. The more appropriate quantum-mechanical expression bears though an important information:

Concept

Electrons in outer shells provide the larger contribution to χ_m .

Diamagnetism in solids is found after summing terms as those in (2.206) for all types of atoms or ions composing the material, so that in this case N indicates the number of atoms of a given type per unit volume. If molecular crystals are considered, (2.206) is again valid: the molecular orbitals are now to be considered and the sum over index i refers to the electrons of single molecules having completely filled shells.

2.7.3 Paramagnetism

Consider now atoms that possess a finite magnetic dipole μ_A , even in the absence of a magnetic field. When the magnetic field is on, two effects occurs. First, dipoles tend to align themselves along the direction of \mathbf{B} in order to lower the magnetic energy: this phenomenon is called paramagnetism. Second, the magnitude of μ_A is modified as in ordinary diamagnetism. In effects, usually diamagnetism accompanies paramagnetism.

The diamagnetic effects being already discussed in the previous section, the attention is now focused on the first effect. The alignment of magnetic dipoles is dependent on the statistical distribution of the initial orientations, and thus on temperature. Thermal averaging within classical statistical distribution yields:

$$\begin{aligned} M &= N\mu_A \frac{\int_0^\pi \cos \theta e^{\beta\mu_A B \cos \theta} \sin \theta d\theta}{\int_0^\pi e^{\beta\mu_A B \cos \theta} \sin \theta d\theta} \\ &= N\mu_A \left(\coth \alpha - \frac{1}{\alpha} \right) \equiv N\mu_A L(\alpha), \end{aligned} \quad (2.207)$$

with $\alpha = \beta\mu_A B$ and $\beta = 1/(k_b T)$. The function $L(\alpha)$ is the Langevin function and dictates the dependence of M on T and B . Since $\alpha \ll 1$ most often, one obtains that

$$M \simeq N\mu_A \frac{\mu_A B}{3k_b T}. \quad (2.208)$$

As a result, the magnetic susceptibility is

$$\chi_m = N \frac{\mu_A^2}{3k_b T}. \quad (2.209)$$

This represents the Curie law for paramagnetic susceptibility, stating that

Concept

In paramagnets, χ_m is small as in diamagnets but positive, and scales as the inverse system temperature.

Turning to the quantum description, one has to revert back to (2.196) for μ_z , so that the magnetization is given by:

$$M = N \mu_b g \frac{\sum_{i=-J}^J M_i e^{M_i g \beta \mu_b B}}{\sum_{i=-J}^J e^{M_i g \beta \mu_b B}}. \quad (2.210)$$

The detailed calculation is now more complicated: it can be demonstrated that

$$\begin{aligned} M &= N \mu_b g J B_J(\alpha), \\ B_J(\alpha) &= \frac{2J+1}{2J} \coth\left(\frac{2J+1}{2J} \alpha\right) - \frac{1}{2J} \coth\left(\frac{1}{2J} \alpha\right), \\ \alpha &= \frac{\mu_b B}{k_b T} g J, \end{aligned} \quad (2.211)$$

in terms of the Brillouin function $B_J(\alpha)$. In the limiting case $\alpha \ll 1$, χ_m turns out to be:

$$\chi_m = N \frac{g^2 \mu_b^2 J(J+1)}{3k_b T}. \quad (2.212)$$

The quantum expression is formally equivalent to the classical one (2.209), provided that μ_A be identified with its quantum analogue: $\mu_A = g \mu_b \sqrt{J(J+1)}$. Therefore, the $1/T$ Curie-law behavior holds in both classical (2.209) and quantum (2.212) paramagnetism.

These results describe well the magnetic susceptibility when the magnetic dipoles are localized on the atoms, ions or molecules that form the material. In solids, atoms are indeed arranged in a crystal and J is not any longer a good quantum number: therefore, χ_m again follows the $1/T$ Curie law $\chi \simeq N g^2 \mu_b^2 S(S+1)/(3k_b T)$, but with the spin quantum number S in place of the total angular momentum J entering the expression of χ_m for one single atom.

When the conduction electrons are considered that behave as free particles, the presence of a magnetic field alters the number of electrons with spin parallel to the field with respect to those with antiparallel spin, so that a finite net magnetic moment appears. This situation, named Pauli paramagnetism, will be discussed in detail in Chap. 4.

Applications. An interesting application of paramagnetism in materials is isentropic cooling, quite useful to cool systems well below 1 K, down to $\approx 10^{-3}$ K. Here is the idea: at fixed temperature, the application of an external magnetic field to a paramagnetic material lowers the entropy, because it essentially induces a higher degree of ordering. In practice, the system is first isothermally magnetized so to lower the entropy, and then it is isentropically (adiabatically) demagnetized. The number of ways to arrange N spins in $2S + 1$ states is $(2S + 1)^N$, so that the spin entropy is $\sigma = k_b \ln[(2S + 1)^N] = Nk_b \ln(2S + 1)$. The spin entropy is thus a function of the population of magnetic states and the latter are functions only of $\mu B / (k_b T)$. Therefore, during the isentropic demagnetization, a constant σ implies constant B/T , so that temperature has to decrease at the rate with which B is reduced. Overall, spins take up entropy from the lattice and the lattice temperature drops down much faster than the spin temperature does.

2.7.4 Ferromagnetism

Diamagnetism and paramagnetism can essentially be described as due to interactions between single induced or permanent magnetic dipoles of atoms and an external magnetic field. All the other atomic magnetic dipoles enter at most to assess statistical averages.

The possibility that magnetic dipoles might interact and be correlated with each other is now brought about. To understand this point, consider the simplest case of two dipoles $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, interacting with an external magnetic field the field \mathbf{B} and among themselves with coupling strength J . The corresponding Hamiltonian is

$$H = -\boldsymbol{\mu}_1 \cdot \mathbf{B} - \boldsymbol{\mu}_2 \cdot \mathbf{B} - J\boldsymbol{\mu}_1 \cdot \boldsymbol{\mu}_2 \quad (2.213)$$

Three different cases may occur: (i) the interaction terms with the external field, that are the first two, dominate; (ii) the dipole-dipole coupling, that is the third term, prevails; (iii) all terms contribute with similar weights, so that they must be considered on the same footing. Case (i) reverts back to the case of diamagnetism or paramagnetism, already discussed. Cases (ii) and (iii) originate ferromagnetism: in essence, the interactions between the magnetic dipoles determine a macroscopic magnetization of the material, that persists even in the absence of an external magnetic field.

The general Hamiltonian form to treat the case (iii) can be usefully cast in the form

$$H = -\frac{2}{\hbar^2} \sum_{i \neq j} J_{i,j} \mathbf{S}_i \cdot \mathbf{S}_j, \quad (2.214)$$

where \mathbf{S}_i and \mathbf{S}_j represent the spins of the given particles i and j , and $J_{i,j}$ their interaction energy. In most situations, the interaction energy washes away with increasing distance between each two particles, so that only the interactions between first neighbors are significant. In the following, the microscopic origin of Hamiltonian (2.214) is discussed, and how this leads to ferromagnetic effects.

Phenomenological model. On phenomenological grounds, in the absence of an external magnetic field the interaction between magnetic momenta originates finite average magnetic dipoles $\langle \mu \rangle$ for each single atom: this then leads to the finite magnetization $M = N\langle \mu \rangle$, scaling with the density N of atoms. The magnetization in turn originates a field B_w

$$B_w = \lambda N \langle \mu \rangle, \quad (2.215)$$

proportional to M via a suitable constant λ . B_w is named Weiss field. If an external field B is switched on, the total magnetic field acting on atoms is

$$B_t = B + \lambda N \langle \mu \rangle. \quad (2.216)$$

The average magnetization due to electrons can thus be expressed as:

$$N \langle \mu \rangle = 2N\mu_b \frac{1}{2} \frac{e^{\beta\mu_b B_t} - e^{-\beta\mu_b B_t}}{e^{\beta\mu_b B_t} + e^{-\beta\mu_b B_t}} = N\mu_b \tanh(\beta\mu_b B_t). \quad (2.217)$$

Repeating for a total spin S the same calculations as those leading to (2.211), one obtains

$$N \langle \mu \rangle = 2N\mu_b S B_S(2\beta\mu_b B_t S), \quad (2.218)$$

in terms of the Brillouin function $B_S(x)$ defined in (2.211). Here, the positions $J = S$ and $g = 2$ have been executed.

In the limiting case $2\beta\mu_b B_t S \ll 1$, one finds that:

$$N \langle \mu \rangle = \frac{4N\mu_b^2 S(S+1)}{3k_b T} (B + \lambda N \langle \mu \rangle)$$

or else

$$N \langle \mu \rangle = \frac{4N\mu_b^2 S(S+1)}{3k_b T} \frac{B}{1 - \frac{4N\lambda\mu_b^2 S(S+1)}{3k_b T}}. \quad (2.219)$$

Defining the temperature

$$T_c = \frac{4N\lambda\mu_b^2 S(S+1)}{3k_b} \quad (2.220)$$

with $\mu \equiv 2\mu_b \sqrt{S(S+1)}$, the magnetization in (2.219) can be cast in the more expressive form

$$M(T) = \frac{N\mu^2 B}{3k_b(T - T_c)}. \quad (2.221)$$

This is the Curie law for ferromagnetism, applying to temperatures $T > T_c$. Notice that the Curie law for diamagnetism and paramagnetism is obtained from (2.221) after setting $T_c = 0$ and considering very small λ values as required by the conditions in case (i) above.

In essence,

Concept

Interactions between magnetic dipoles may lead to macroscopic magnetic ordering below a critical temperature T_c given by (2.220). This magnetic ordering spontaneously takes place in the absence of an external magnetic field, since it corresponds to a ground state with lower energy than the non-ordered one .

Starting from the finite $M(T = 0)$ value, the magnetization decreases with a $T^{3/2}$ law while temperature increases up to T_c . The meaning of T_c is better understood by looking at the solutions to (2.217) and (2.218) when $B = 0$. Indeed, one finds that the solution $M = 0$ is possible only for $T > T_c$: thus, above the critical temperature the magnetization vanishes. This behavior defines a transition from a normal state with $M = 0$ to a ferromagnetic state with $M \neq 0$.

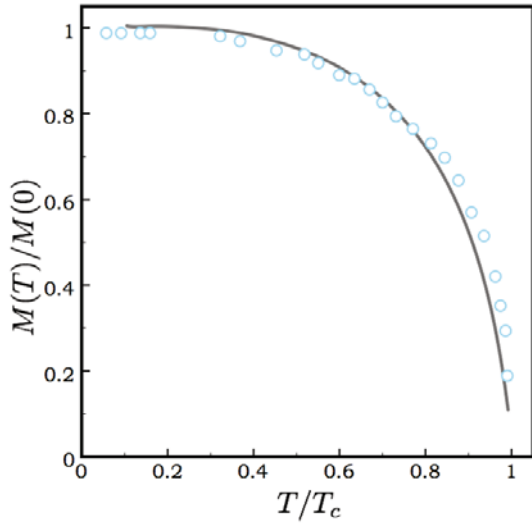
This is evident in Fig. 2.41, where experimental data of the magnetization as a function of temperature are displayed for a Nickel sample compound, along with the theoretical prediction from (2.218) for the case with $S = 1/2$. Since Ni has 8 electrons in d states, one would have $S = 1$. The fitted value $S = 1/2$ suggests that s and d bands somehow overlap. In general, one finds that at $T = 0$ the magnetization can be cast in the form $M(0) = n_b \mu_b N$, with n_b representing a sort of effective magneton number. Measurements are usually consistent with non-integral values of n_b : this can be due to the effect of spin-orbit interactions, or to a local magnetization induced around conduction electrons as in Pauli paramagnetism, or else because the size of the aligned magnetic moments is not uniform (see below). In the case of Ni with $T_c = 627$ K, one extracts the value $n_b \simeq 0.6$ per atom. For Fe with $T_c = 1043$ K: $n_b \simeq 2.22$. For Co with $T_c = 1388$ K: $n_b \simeq 1.72$. For Gd with $T_c = 292$ K: $n_b = 7.63$, and so on.

To summarize, for $T > T_c$ the susceptibility behaves as $\chi_m = N\mu^2 / [3k_b(T - T_c)]$. For $T < T_c$ instead, a macroscopic magnetization manifests even at zero magnetic field. T_c signals the temperature below which the transition to the new ground state with ferromagnetic order occurs. Critical temperatures T_c are often higher than room temperature, so that ferromagnets are easily available for everyday life applications.

Quantum-mechanical model. The classical interaction between magnetic dipoles is too small to justify such a gigantic effect on the magnetization. A quantum approach is better suited here, where exchange interactions can be introduced and play a role. In order to make the treatment easier, consider two atoms A and B, each with one electron characterized by wavefunctions $\psi_A(\mathbf{r}_1)$ and $\psi_B(\mathbf{r}_2)$, respectively. The positions \mathbf{r}_1 and \mathbf{r}_2 are considered to be centered on atoms A and B, respectively. The system Hamiltonian has the form $H = H(1) + H(2) + H(1, 2)$, with $H(1)$ and $H(2)$ the Hamiltonians for the two isolated atoms A and B, whose eigenfunctions are $\psi_a(\mathbf{r}_1)$ and $\psi_b(\mathbf{r}_2)$. The interaction term $H(1, 2)$ can be cast in the form

$$H(1, 2) = -\frac{e^2}{|\mathbf{r}_2 - \mathbf{r}_A|} - \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_B|} + \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} + \frac{e^2}{|\mathbf{r}_B - \mathbf{r}_A|}. \quad (2.222)$$

Fig. 2.41 Occurrence of ferromagnetic effects. The magnetization $M(T)$ scaled to the value $M(0)$ at zero temperature vs. temperature T scaled to T_c . The case of Ni is shown. Symbols: experimental data [26]. Solid curve: theoretical prediction resulting from a fit to (2.218) with $S = 1/2$



Here, the first (second) term represents the interaction of the electron of atom B (A) with the nucleus of atom A (B), whereas the third (fourth) term represents the repulsive interaction between the two electrons (two nuclei). It is known that the total wavefunctions of the system, namely Ψ_1 and Ψ_2 , are set up by considering the symmetric (antisymmetric) space wavefunction times an antisymmetric (symmetric) spin wavefunction.

The good quantum numbers for such wavefunctions are related to the total spin \mathbf{S} , total spin squared \mathbf{S}^2 , and single-particle spin squared \mathbf{S}_1^2 and \mathbf{S}_2^2 quantum operators. The eigenvalues of \mathbf{S}^2 corresponding to Ψ_1 and Ψ_2 are 0 and $2\hbar^2$, respectively. The energies of the system can be calculated, that are:

$$\begin{aligned}\langle \Psi_1 | H | \Psi_1 \rangle &= E_A + E_B + J, \\ \langle \Psi_2 | H | \Psi_2 \rangle &= E_A + E_B - J,\end{aligned}\tag{2.223}$$

where E_A and E_B the electronic energies of atoms A and B, respectively. As to J , this is given by the matrix element

$$J = \int d\mathbf{r}_1 d\mathbf{r}_2 \psi_A^*(\mathbf{r}_1) \psi_B^*(\mathbf{r}_2) H(1,2) \psi_A(\mathbf{r}_2) \psi_B(\mathbf{r}_1).\tag{2.224}$$

The sign of J determines the ground state wavefunction. If $J < 0$, the ground state is represented by Ψ_1 with eigenvalue 0 for \mathbf{S}^2 . If $J > 0$ instead, the ground state is represented by Ψ_2 with eigenvalue $2\hbar^2$ for \mathbf{S}^2 . On the other hand, one has

$$\mathbf{S}^2 = \mathbf{S}_1^2 + \mathbf{S}_2^2 + 2\mathbf{S}_1 \cdot \mathbf{S}_2,\tag{2.225}$$

so that the two results in (2.223) can be easily cast in the expressive form

$$\begin{aligned} E &= E_A + E_B + J \left(1 - \frac{\mathbf{S}^2}{\hbar^2} \right) \\ &= E_A + E_B - \frac{J}{2} - 2 \frac{J}{\hbar^2} \mathbf{S}_1 \cdot \mathbf{S}_2. \end{aligned} \quad (2.226)$$

The last term on the right-hand side of (2.226) is in the same form as (2.214): this simple model supports the idea that ferromagnetic effects originate from exchange terms between the electrons on neighboring atoms. $H(1,2)$ contains two negative and two positive terms, so that the overall sign of J depends on their competition. It can be shown that J is negative when the atoms are very close to each other, becomes positive when the distance increases, and tends to zero at further larger distance. Hamiltonian (2.214) is called Heisenberg Hamiltonian, from which magnetization, critical temperature, specific heat, and excitations can be calculated by means of analytical and numerical methods.

2.7.5 Antiferromagnetism and other magnetic orderings

A few relevant concepts emerge indeed from the analysis of (2.214). In order to make the treatment easier, the model-case of an infinite linear chain is considered in the rest of the section. In \hbar units, that is performing the substitution $\mathbf{S}_i \rightarrow \hbar \mathbf{S}_i$, the Heisenberg Hamiltonian for an infinite linear chain can be cast in the form

$$H = -2J \sum_n \mathbf{S}_n \cdot \mathbf{S}_{n+1}. \quad (2.227)$$

First, one notices that Hamiltonian (2.227) may lead to two different orderings. If the spins are aligned all in the same direction, the energy density is

$$U_0 = -2JNS(S+1), \quad (2.228)$$

with S the eigenvalue of \mathbf{S}_n^2 . If only one spin is flipped over, the energy density turns into:

$$U_1 = U_0 + 8JS(S+1). \quad (2.229)$$

Therefore, one has $U_0 < U_1$ when $J > 0$, and $U_1 < U_0$ when $J < 0$. When $J > 0$, the state with lower energy corresponds to the situation where all the spins are aligned: this is indeed the ferromagnetic state, depicted in Fig. 2.42 (a). When $J < 0$ though, it can be shown that the state is favored in which adjacent spins are ordered with alternating orientations. This is schematically pictured in Fig. 2.42 (b). In canted antiferromagnetic ordering instead, adjacent spins are alternating but at intermediate angle, as sketched in Fig. 2.42 (c).

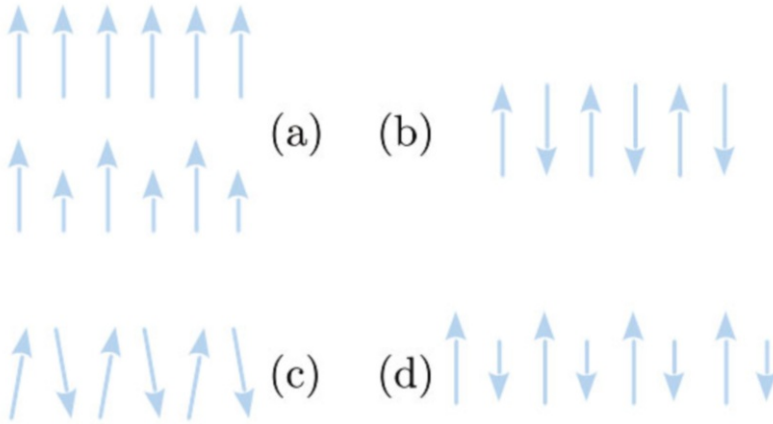


Fig. 2.42 Schematic representation of different magnetic orderings: (a) ferromagnetic (top) and ferrimagnetic (bottom), (b) antiferromagnetic, (c) canted antiferromagnetic, and (d) antiferromagnetic order. Each arrow represents a spin at a lattice site. In order to catch the basic concept, a 1D picture is provided

In essence,

Concept

When $J < 0$, the lowest-energy (ground) state corresponds to an ordering where two spin lattices exist, whose magnetic moments have equal sizes but the corresponding spins are antiparallel to each other. This new ground state is called antiferromagnetic state.

The transition temperature for antiferromagnetic ordering is named Néel temperature T_N . It can be shown that $T_N = C\mu$ is proportional to magnetic moment μ . The antiferromagnetic susceptibility for temperatures $T > T_N$ can be shown to behave as $\chi_m = 2C/(T + \theta)$, where θ is a constant depending on the material. Below T_N , the antiferromagnetic susceptibility rapidly drops to zero if the field is parallel to the spin orientation, and tends to the constant $1/\mu$ value in the perpendicular case. Antiferromagnetic materials are for example crystals of Cr with $T_N = 308$ K; manganese compounds like MnO with $T_N = 116$ K, MnS with $T_N = 160$ K, MnTe with $T_N = 307$ K, MnF₂ with $T_N = 67$ K; iron compounds such as FeO with $T_N = 198$ K, FeF₂ with $T_N = 79$ K; cobalt CoO with $T_N = 291$ K and nickel NiO with $T_N = 525$ K compounds, and so on.

Other ground states may set in when the size of magnetic moments of two spin lattices is different.

Concept

Ferrimagnetic ordering occurs when the spin orientations of two sublattices with different net magnetic moments are parallel. Antiferrimagnetic ordering occurs when the spin orientations of two sublattices with different magnetic moments are antiparallel.

Ferrimagnetic order is depicted in the bottom panel of Fig. 2.42 (a) and antiferrimagnetic order in Fig. 2.42 (b). A popular ferrimagnet is for example magnetite, Fe_3O_4 or $\text{FeO}\cdot\text{Fe}_2\text{O}_3$. In general, magnetic oxides with formula $\text{MO}\cdot\text{Fe}_3\text{O}_4$ are named ferrites, with M being a divalent cation such Zn, Cd, Fe, Ni, Cu, Co, or Mg. In magnetite for example, the three-valent Fe^{+3} ferric ions have $S = 5/2$ and $L = 0$, thus contributing with $5\mu_b$ to $M(0)$; the divalent Fe^{+2} ferrous ions have $S = 2$ and should contribute with $4\mu_b$, besides any L -like term. Thus, if all the spins were parallel, the effective number of magnetons per formula Fe_3O_4 would be $n_b = 2 \cdot 5 + 4 = 14$. The value inferred from experiments is instead $n_b \simeq 4$, that can be explained if one assumes that the moments of ferric ions in tetrahedral and octahedral sites are antiparallel and cancel out, leaving only the moments of the ferrous ions on the octahedral sites.

In fact, whatever the ordering be, specimen are often composed of smaller regions distinguished by the direction of the macroscopic magnetization.

Concept

The formation of such domains originates from the competition between different types of involved energies, that are due to exchange, anisotropic and magnetic effects: domains set in because this is what makes the overall energy lower. Domains are separated by a sort of walls, named Bloch walls, whose thickness and energy are dictated by the exchange coupling energy J and the anisotropy energy K : in particular thickness is found to scale as $a\sqrt{J/K}$ and energy per unit area as \sqrt{JK}/\sqrt{a} in terms of the lattice constant a .

2.7.6 Structure of excitations in ferromagnets and antiferromagnets

After having explored the ground states corresponding to the different magnetic orderings, the question arises on how are the lowest-energy excitations characterized. In order to catch the basic features, calculations are performed for just the ferromagnetic state. Similar calculations can be performed for antiferromagnets, for which the results are given. The introduction of the magnetic dipole $\boldsymbol{\mu}_n = 2\mu_b\mathbf{S}_n$ casts the energy in the form:

$$U = -\frac{1}{2} \sum_n \boldsymbol{\mu}_n \cdot \mathbf{B}_n, \quad (2.230)$$

with

$$\mathbf{B}_n = \frac{J}{\mu_b} (\mathbf{S}_{n-1} + \mathbf{S}_{n+1}). \quad (2.231)$$

In this form, (2.230) can be handled as if it were a classical interaction term. In order to determine the excitations, the equation of motion for the angular momentum $\hbar \mathbf{S}_n$ can be a convenient starting point, that is:

$$\hbar \frac{d\mathbf{S}_n}{dt} = \boldsymbol{\mu}_n \wedge \mathbf{B}_n = 2JS_n \wedge (\mathbf{S}_{n-1} + \mathbf{S}_{n+1}). \quad (2.232)$$

Expression (2.232) is a set of nonlinear equations for the given single spin. Since the interest is on the excitations with lowest energy, a linear expansion about the ground state can be performed. In fact, after setting $\mathbf{S}_n = \mathbf{S}_z + \boldsymbol{\sigma}_n$, linearization of (2.232) with respect to $\boldsymbol{\sigma}_n$ yields:

$$\begin{aligned} \left(\frac{d\boldsymbol{\sigma}_n}{dt} \right)_y &= \frac{2J}{\hbar} S_z (\boldsymbol{\sigma}_{n-1} + \boldsymbol{\sigma}_{n+1} - 2\boldsymbol{\sigma}_n)_x, \\ \left(\frac{d\boldsymbol{\sigma}_n}{dt} \right)_x &= \frac{2J}{\hbar} S_z (\boldsymbol{\sigma}_{n-1} + \boldsymbol{\sigma}_{n+1} - 2\boldsymbol{\sigma}_n)_y. \end{aligned} \quad (2.233)$$

The quantity $\boldsymbol{\sigma}^+ = \sigma_x + i\sigma_y$ satisfies the time-dependent equations

$$i \frac{d\boldsymbol{\sigma}_n^+}{dt} = \frac{2J}{\hbar} S_z (\boldsymbol{\sigma}_{n-1}^+ + \boldsymbol{\sigma}_{n+1}^+ - 2\boldsymbol{\sigma}_n^+), \quad (2.234)$$

that are formally equivalent to those considered when studying the electronic band structure of infinite linear chains. Similar to that case, the solution to (2.234) can be searched of the form

$$\boldsymbol{\sigma}_n^+ = A e^{i(kna - \omega t)}, \quad (2.235)$$

with a the lattice constant. Inserting (2.235) into (2.234), expression (2.234) is satisfied provided that

$$\hbar \omega(k) = 4JS_z (1 - \cos ka), \quad (2.236)$$

which provides the spectrum of the lowest-energy excitations. Thus, the lowest-lying ferromagnetic excitations of the spin \mathbf{S}_n in $x-y$ plane, have a cos-like dispersion relation. Though obtained within a classical treatment, this result remains valid also in quantum-mechanical sense. In essence,

Concept

The lowest-energy normal modes of an infinite linear ferromagnetic spin chain are spin waves, characterized by a quadratic dispersion at long wavelengths or else vanishing wavevectors $\mathbf{k} \rightarrow 0$. These excitations are called magnons.

As usual in quantum mechanics, magnons can be intended to be either spin waves or particles. A pictorial representation of these waves is reproduced in the top panel of Fig. 2.43. A magnon effective mass can be defined and easily calculated from (2.236), yielding $m^* = \hbar^2 / (4JS_z a^2)$. In the long wavelength limit $k \rightarrow 0$, the dispersion relation $\hbar \omega(k) = \hbar^2 k^2 / (2m^*)$ is quadratic in k . The bottom panel of Fig. 2.43

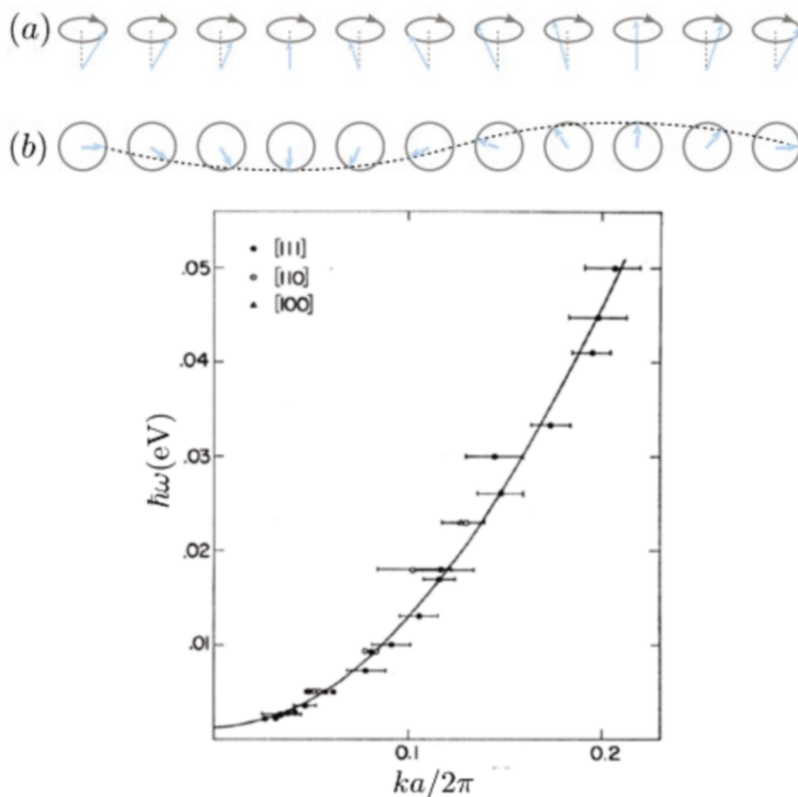


Fig. 2.43 Lowest-energy excitations from a ferromagnetic ground state. Top panel: sketch of a spin wave in a linear chain from front (a) and top (b) viewpoints. Bottom panel: dispersion relation for the energy of spin waves in a Fe-Co alloy. Symbols with error bars: experimental data from inelastic neutron diffraction [27]. Solid curve: theoretical fit, showing that the quadratic dependence is valid in the long-wavelength $k \rightarrow 0$ regime. Reprinted with permission from [27]. Copyright (1960) by the American Physical Society

displays the experimental data on the dispersion of spin waves in a Fe-Co alloy, as measured from inelastic neutron diffraction [27]. The fit to the data shows the validity of the predicted quadratic dependence in the long-wavelength $k \rightarrow 0$ regime.

Similar calculations can be carried out for the remaining ordered states, antiferromagnetic or (anti)ferrimagnetic, taking care of writing down separate equations for the two sublattices and assigning in each case the correct sign to the J coupling. This coupling between sublattices modes leads to a remarkable result:

Concept

In antiferromagnets, the long wavelength dispersion relation for the lowest-energy excitations is of the type $\hbar\omega(k) \simeq Jak$, that is linear in k .

Quick Questions

Q15. Magnetic tapes for recording are thin strips of plastic, coated on one side with a film of tiny permanent magnetic particles. Strong magnetic fields are produced by a tiny electromagnet in the form of a ring, that is a ring with a coil of conducting wire wounding around it. The ring is cut in correspondence of the point where it touches the magnetic tapes moving underneath: the opposite ends of this cut work as north and south magnetic poles. During recording, mechanical sound waves are converted by a transducer into electric alternating currents in the winding wire, and therefore into fluctuating magnetic fields in the ring's gap: these magnetize the tape, leaving a magnetic pattern that can be afterward played back in a similar but reversed process. Before the recording, the tape is accurately erased by a second tape head operating with large, high-frequency currents that are able to provide random orientation to the magnetic moments of the tape [1]. All this considered, would you better use Gd or Fe to coat the plastic strip?

Answer. Certainly Fe. At variance with Fe, Gd has a low critical temperature $T_c = 292\text{ K}$ for the transition to ferromagnetic state. Thus, higher would be the risk of demagnetization. Indeed, best tapes are composed of iron or various kinds of ferrites. Notice that magnetic data storage in computer hard disks uses concepts similar to tape recording, but: the recording is digitalized in 0 and 1, thereby requiring coatings with micrometer thickness; digital information is added to help locating and retrieving the bit, while the reading head shifts all the way from the disk edge to the center and viceversa; additional electronic and optical feedback mechanisms have been introduced to find the recorded bits, while the size of the hard disks and thus of the stored bits is progressively miniaturized.

Q16. Magnetic levitation is based on the concept that like magnetic poles repel each other. However, realizing magnetic levitation with objects possessing permanent magnets on a permanently magnetic surface is not practical: while the object stays on top of the surface, whatever tiny lateral displacement would make the object flipping and falling down. This is indeed a situation of unstable equilibrium. For this reason, feedback mechanisms are introduced that make the equilibrium dynamically stable. One possibility is to use the concept of electromagnetic induction and use electromagnets as those that make working generators of alternate currents and, by the reverse process, motors. In electromagnets, the intensity and direction of the coil currents can be easily adjusted to produce the wanted magnetic fields. Levitating trains are however based on a more practical scheme, where electromagnets are mounted on the train, that runs on just metallic rails [1]. Discuss how such a scheme might work.

Answer. While moving, the electromagnets mounted on the train produce a variable magnetic field, that acts on the rail. Therefore, on the metallic rail a current appears by magnetic induction, which in turn makes the rail magnetic. By Lenz law, the magnetic rail produced by the rail opposes to that originating it, and the train levitates. Unfortunately, currents in the back of the train tend to decay and are thus smaller than the currents in the front of the train, where they have just been created. Therefore, magnetic repulsion is stronger in the front than in the back, leading to a magnetic drag force opposing to the train motion. Of course, the faster is the train, the smaller is the differential effect. The magnetic resistance is typically the largest around the speed of 30 Km/h, that is the least speed to start levitation. For this reason, levitating trains are equipped with retractile wheels to be used during take off and landing [1].

2.8 Pair-correlated ground states and excitations: the case of superconductors and superfluids

2.8.1 Superfluidity, superconductivity, Bose Einstein Condensation

The electronic structures considered in the first part of the present Chapter are based on the idea that the ground state originates from a combination of single-particle electronic orbitals, whatever complicated can be the ways in which interactions modify them. This is one out of many different possible types of ground states.

⁴He bosonic superfluidity. At the beginning of XX century, Kamerlingh Onnes found a method to liquefy the boson isotope of Helium, ⁴He, so that temperatures below about 5 K could be reached. Very soon, it was found that below a critical temperature $T_c \simeq 2.2$ K, liquid ⁴He began to expand rather than contract and flowed without dissipation in pipes, provided that the linear velocity of flow be smaller than a critical v_c . If put in a rotating bucket and depending on temperature below T_c , the fluid shows to possess irrotational character: below a given critical angular velocity ω_{c1} , the fluid does not follow the rotating walls of the container, while above ω_{c1} and below ω_{c2} the formation of vortexes sets in as energetically favorable, that is density singularities around which the liquid can rotate. Finally, above ω_{c2} the whole fluid normally rotates with the bucket [41]. Above the threshold given by the linear v_c and angular ω_{c2} velocities then, as in ordinary liquids the fluid slows down because it is dragged along by the walls. The occurrence of these overall phenomena has been given the name of superfluidity. A first indication for the critical velocity can be inferred after an argument due to Landau, as follows. Dissipation enters via the progressive excitation of internal motions with energy $\varepsilon(\hbar\mathbf{k})$ and momentum $\hbar\mathbf{k}$ in the fluid at rest. In the fluid moving with velocity \mathbf{v} however, the excitation energy reads $\varepsilon(\hbar\mathbf{k}) + \hbar\mathbf{k} \cdot \mathbf{v}$. When this is negative, the energy of the liquid is being decreased. Since the minimum value corresponds to the case in which \mathbf{k} is parallel

and opposite to \mathbf{v} , viscous flow sets in when $v > \varepsilon(\hbar k)/(\hbar k)$. Since long-wavelength density fluctuations are sound waves, with energy $\propto c_s k$, that is, proportional to k via the speed of sound c_s , the critical velocity would precisely be c_s . However, it is seen that the critical velocity is up to two orders of magnitude smaller than the speed of sound. In fact, it was argued that the normal fluid could be regarded as a dilute gas of weakly-interacting excitations obeying Bose statistics. Besides sound waves at $k \rightarrow 0$, a second type of excitation was inferred and demonstrated to exist, named rotons, that according to Feynman are tiny vortex rings, so tiny that only one atom can pass through its center. The least energy Δ required to excite a roton is finite and corresponds to a minimum gap of the excitation spectrum of the liquid at a finite wavevector k_R : that is, the excitation spectrum linearly goes up in the $k \rightarrow 0$ limit, then reaches a maximum in correspondence of the wavevector dictated by the excitation Brillouin-zone edge, then reaches a minimum that is at energy Δ above zero, and finally proceeds towards the spectrum of free particles as $k \rightarrow \infty$ [51]. Overall,

Concept

The definition of superfluidity can be related to the resistance that a fluid offers to be twisted by transverse, shear external disturbances and, in second place, to longitudinal compressional perturbations .

Bose-Einstein Condensation. The thermal properties of superfluid He leads to the idea that the system behaves as if it carries no entropy, like it occurs for one single quantum state. In fact, since the beginning

Concept

Superfluidity has been considered a phenomenon close to Bose-Einstein Condensation.

Bose-Einstein Condensation (BEC) phenomenon had been predicted by Bose for photons and extended by Einstein to non-interacting atoms with bosonic character. In the form predicted by Bose and Einstein, BEC has been observed in 1995 by the groups of Cornell and Wiemann at JILA, and of Ketterle at MIT in quantum gases of bosonic ^{87}Rb and ^{23}Na atoms, after magnetically trapping something like $10^4 - 10^5$ of them and cooling them down below ≈ 170 nK temperatures by combined techniques of laser and evaporative cooling [28]. Appendix 2.13 contains a more detailed discussion of this phenomenon along with its more recent realizations.

Concept

In Bose-Einstein condensation, a macroscopic occupation of the lowest energy state occurs below a critical temperature, leading to a sort of condensation but in momentum or energy space, instead than in real space like in a vapor-to-liquid

phase transition. As the temperature drops, the De Broglie wavelength of the particles becomes infinitely large. That is to say, the kinetic energy of the lowest state drops to zero and the particle wavefunctions hook together extending over all the volume, so that the particles have large probability to be at any point in space. Under such conditions, the system behaves as if it were one single particle with well defined momentum and energy.

Thus, the argument connecting superfluidity to Bose-Einstein Condensation is that slowing down the motion of condensed particles requires a coupling to all or most of all of them at the same time. This in turn would require a large energy, larger indeed than that for the occurrence of either sound waves, or rotons, or vortexes.

Bose-Einstein Condensation is a direct consequence of the Bose-Einstein statistics, where the number of particles N distributed at temperature T and with density of states $g(\varepsilon)$, is given by

$$N(T, \mu) = V \int_0^\infty d\varepsilon g(\varepsilon) \frac{1}{e^{\beta(\varepsilon - \mu)} - 1}, \quad (2.237)$$

and $\beta \equiv (k_B T)^{-1}$. While temperature goes down, conservation of the number of particles requires that the chemical potential increases from negative values to 0. At the temperature T_c such that $\mu = 0$, the quantity $N'(T) = N(T, \mu = 0) < N$ and conservation of the particle number can be accomplished only if a macroscopic occupation N_0 of particles in the ground state occurs, so that $N(T) = N_0(T) + N'(T)$.

The connection between Bose-Einstein condensation and superfluidity is however not so immediate, since experiments have shown that only a fraction $N_0/N \approx 10\%$ of liquid Helium atoms typically belongs to the condensate while the system is entirely superfluid: thus,

Concept

The density of Bose-Einstein condensed particles differ from the density of superfluid. This is an effect of interactions between the particles: they introduce a finite probability that non-condensed particles be carried along with the condensate during the superfluid flow.

³He fermionic superfluidity. In addition, similar superfluid behavior has been shown in ³He, that at variance with bosonic ⁴He has a fermionic character. If a connection between superfluidity and BEC exists, besides the difference between condensate and superfluid density, one is left to explain how

Concept

Fermions as well can undergo a form of condensation in momentum space, by occupying all one and the same ground state while the Pauli exclusion principle would prevent single fermions to do so.

Superconductivity in metallic and in ceramic compounds. Soon after He liquefaction, normal metals cooled at temperatures below 10-20 K have as well been shown to possess a behavior similar to superfluid Helium, though this time the transport of electric current is involved [29] instead than a neutral fluid flow. In Pb for example, it has been shown that below the critical temperature $T_c \simeq 7$ K, the resistivity vanishes under the action of an external electric field, as if any dissipation mechanism on electrons motion were frozen out. More than three decades later Bardeen, Cooper, and Schrieffer published the microscopic theory of superconductivity, giving full account of the experimental facts [30].

In 1986 and later on, a similar behavior of electrical resistivity has been found [31] in completely different types of materials, rare-earth based ceramics indeed, with critical temperatures as high as 150 K. The first discovered material with such characteristics was a non-stoichiometric compound $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$, that is oxygen-doped: its stoichiometric version $\text{YBa}_2\text{Cu}_3\text{O}_6$ is a superconductor with $T_c = 93$ K, T_c being in fact dependent on doping. The crystal structure of all these new superconductors is quite complex indeed. In essence, they are composed by planes of CuO_2 where the atoms Cu form a square lattice surrounded by four O atoms; adjacent planes of CuO_2 are separated by insulating chains. As a general rule, T_c increases by increasing the number of planes in each unitary cell with a well defined best maximum value. This is illustrated in Fig. 2.44.

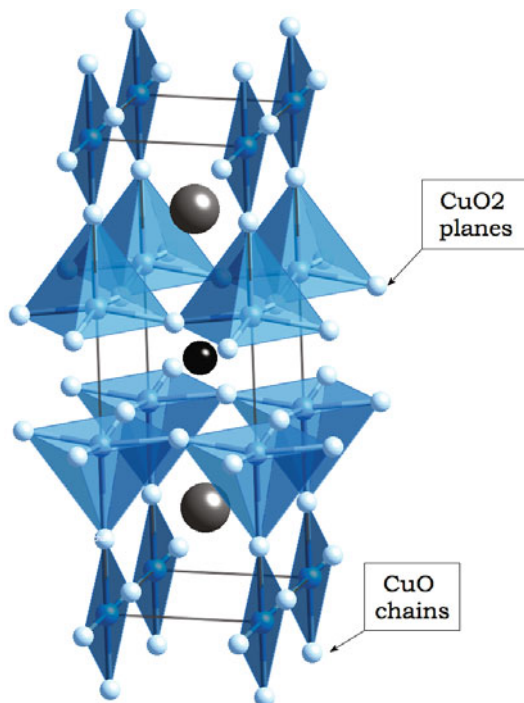
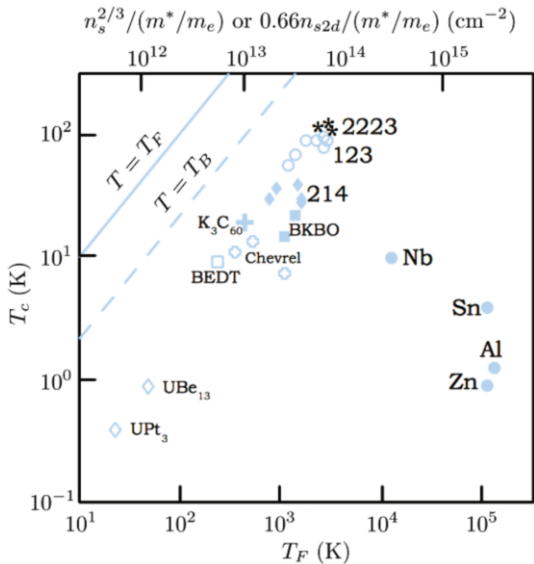


Fig. 2.44 Crystal structure of high- T_c superconductors, based on CuO_2 layers separated by insulating chains

Fig. 2.45 Superconductivity and superfluidity: the universal Uemura plot. The critical temperatures T_c for a number of superconducting and superfluid materials vs. their Fermi temperature T_F (lower axis) or the temperature T_B corresponding to Bose-Einstein condensation of fermion pairs (upper axis) [34]



The question is therefore whether all these phenomena are related to each other as their phenomenology seems to suggest. On one side, non-viscous flow in superfluids and in superconductors, on the other side, superfluidity and Bose-Einstein Condensation. If this were the case, the question would arise on which type of ground state admits such peculiar effects, that is able to comprehend so different systems. The assessment of the first question becomes more clear after the discovery of high- T_c superconductors with the famous Uemura plot, reported in Fig. 2.45. This depicts in log-log scale the critical temperature T_c for a number of different superconductors and superfluids versus the Fermi temperature $T_F \equiv E_F/k_b$ or an equivalent critical temperature T_B for Bose-Einstein Condensation (BEC) of pairs of fermions. Most of the materials lay down within a universal line staying below the T_B line. The BEC realizations in quantum gases of bosonic ^{87}Rb and ^{23}Na are consistent with the Uemura plot. So does also the analog realization of fermionic superfluidity in confined quantum gases of fermionic ^{40}K atoms cooled down to ≈ 50 nK temperatures [33], as predicted [32].

In essence, the Uemura plot extended to the more recent BEC experiments with Fermi and Bose quantum gases, supports the key idea that

Concept

Superconductivity or superfluidity of both bosonic and fermionic-like systems are related phenomena.

The underlying concept that superconductivity or superfluidity is Bose-Einstein Condensation [41]. A BEC system is characterized by a wavefunction $\psi(\mathbf{r}) =$

$\sqrt{n_c(\mathbf{r})}e^{i\phi(\mathbf{r})}$ with a condensate density $n_c(\mathbf{r})$ and a global space-dependent phase $\phi(\mathbf{r})$. Two consequences are in order. First, a macroscopic fluid motion may set in: if $n(\mathbf{r})$ is lowly varying in the space, the quantum-mechanical condensate current

$$\mathbf{j}_c \equiv \frac{1}{(2mi)} [\psi(\mathbf{r})^* \nabla_{\mathbf{r}} \psi(\mathbf{r}) - \psi(\mathbf{r}) \nabla_{\mathbf{r}} \psi(\mathbf{r})^*] = \frac{n_c(\mathbf{r})}{m} \nabla_{\mathbf{r}} \phi(\mathbf{r})$$

and thus the superfluid velocity \mathbf{v}_s is dictated by the space gradient of the phase. The superfluid velocity is therefore a macroscopic variable: unlike a normal liquid, a superfluid can be at equilibrium even when flow exists with velocity \mathbf{v}_s relative to the pipe walls. Second, it can be anticipated that for a fermion system the condensate wavefunction $\psi(\mathbf{r})$ is to be substituted by a pairing wavefunction $\psi(\mathbf{r}) \rightarrow p(\mathbf{r}) \equiv \langle \psi_{\downarrow}(\mathbf{r}) \psi_{\uparrow}(\mathbf{r}) \rangle$ resulting from a statistical average $\langle \dots \rangle$ within the whole system, and corresponding to the situation in which the Bose particle (integer spin) is represented by two paired fermions with opposite spins. Notice that in general fermions can be also paired with angular momenta higher than 0, as in $(\downarrow\uparrow)$, and at variance with the true boson wavefunctions, fermion-pairs wavefunctions do spatially overlap.

In the following, the focus is placed on the above key idea that superconductivity or superfluidity of both bosonic and fermionic-like systems are related phenomena. Considered that focus of this textbook is on solid-state systems, the mainstream is here taken on superconductors, bringing in the other systems whenever convenient. The physics is discussed at a phenomenological level first, and then within the most successful theoretical framework, that is the Bardeen, Cooper and Schrieffer (BCS) theory. The BCS theory contains an essential ingredient, that is the possibility that carriers may pair into a new type of ground state. For normal metals becoming conventional superconductors, the microscopic pairing mechanism has been clearly demonstrated to be mediated by coupling with the vibrational states of the crystal investigated in Chap. 3. No accepted microscopic theory is available today for high- T_c superconductors instead. However,

Concept

Whatever the microscopic mechanism might be for the pairing, the BCS theory has shown to be first of all a general framework to treat superconductivity and superfluidity phenomena involving fermionic-like particles. It can actually be demonstrated that BEC theory for fermionic systems can be made to continuously evolve towards the theory of Bose-Einstein Condensation for bosonic-like particles, after varying appropriate parameters [43, 44].

2.8.2 Phenomenology of superconductors

In this section, the essential experimental facts are illustrated in regard to metallic superconducting compounds.

Electrical conductivity, radiation absorption and specific heat. The most straightforward manifestations of superconductivity in normal metals are electric conductivities as large as $10^{25} \Omega^{-1}\text{cm}^{-1}$ below T_c , or else a practically zero resistivity. This behavior is displayed in Fig. 2.46 (a).

Normal superconductors do not absorb radiation with frequency smaller than a threshold ω_c of the order of $\hbar\omega_c \simeq 4k_bT_c$ at $T = 0$. At frequencies $\omega > \omega_c$ the material behaves as a normal metal. This behavior is displayed in Fig. 2.46 (b).

The specific heat below T_c behaves in a peculiar way, as if the number of particles that can be thermally excited were an exponentially increasing function of temperature while T approaches T_c from below: in particular, $c_V \propto e^{-\Delta/(k_bT)}$, the values increasing while T approaches T_c from below. Above T_c , the specific heat is found to be $c_V \propto T$ as in normal metals. This behavior is displayed in Fig. 2.46 (c).

Temperature-dependent energy gap. All these facts are consistent with a structure of the ground-state energy characterized by a finite energy gap Δ between the ground and the first excited states. In the presence of such a gap, scattering processes determining the material resistivity are highly depressed: no final states are available for processes in which electrons change their momentum and energy if initial and final states are separated by a gap large enough, that is larger than thermal energy k_bT . Optical transitions are also highly depressed below T_c unless the energy of the quantum of light exceeds the gap in the electronic energy spectrum. The presence of a gap also motivates on general grounds the exponential behavior of the specific heat.

Of course, the quantity Δ appearing in the specific heat must be a vanishing function of T while $T \rightarrow T_c$. The measured temperature dependence of the gap energy is consistent with

$$\Delta(T) = \Delta_0 \sqrt{1 - \frac{T}{T_c}} \quad (2.238)$$

where $\Delta_0 \equiv \Delta(T = 0)$ is the gap at $T = 0$. This is displayed in Fig. 2.47 (a). In addition, data on the low-temperature gap Δ_0 and T_c are compatible with the relation

$$\frac{2\Delta_0}{(k_bT_c)} \simeq 3.5, \quad (2.239)$$

holding for most superconductors.

Behavior in magnetic fields. Additional experimental facts are relevant to understanding the origin of superconductivity. In 1933 it was discovered that below T_c magnetic fields $H < H_c$ below a critical threshold H_c were efficiently expelled by the superconductor: in this effect, named after Meissner and Ochsenfeld, the material behaves as if it were a perfect diamagnet. Above H_c , the material recovers its normal behavior. The critical field changes with temperature according to the law

$$H_c(T) = H_0 \left[1 - \left(\frac{T}{T_c} \right)^2 \right] \quad (2.240)$$

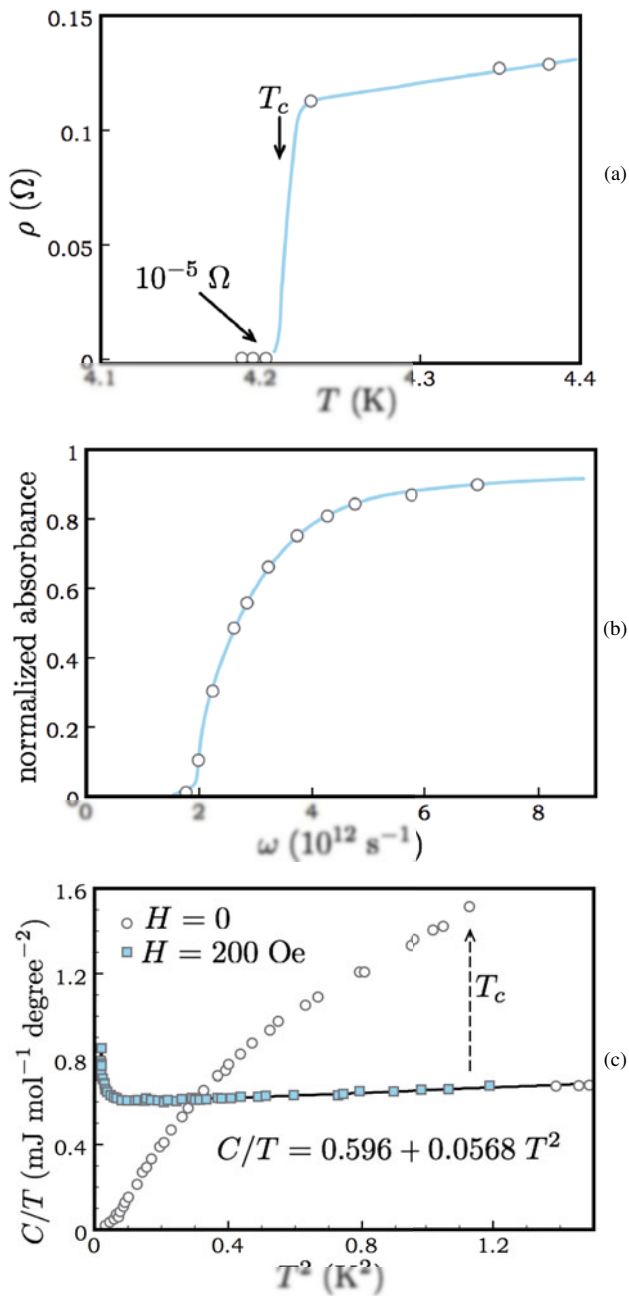


Fig. 2.46 Phenomenology of metallic superconductors. (a) Resistivity of Hg vs temperature T [35]. (b) Light absorption of In vs. the frequency ω of the incident light: notice that the threshold for vanishing absorption corresponds to the condition $\hbar\omega < 4k_B T_c$ [36]. (c) Specific heat of Ga vs. temperature T . Open symbols: superconducting state. Filled symbols: normal state obtained after applying a magnetic field above critical threshold for destroying superconductivity. To help the reading of the scaling laws, actually C/T is displayed vs. T^2 [37]

where $H_0 \equiv H_c(T=0)$ is the critical field at $T=0$. This is displayed in Fig. 2.47 (b). If the magnetic induction field $B=0$ has to vanish in the superconductor bulk, then it must be $H+4\pi M=0$, with M the magnetization of the material: the superconductor creates electric currents that completely screen the field B . These are the so-called type-I superconductors. More generally, it is also possible that a range of magnetic fields H exist with $H_{c1} < H < H_{c2}$, for which the field lines of B are only partially expelled from the material. This occurs while at the same time the resistivity is zero. For fields H larger than H_{c2} instead, the superconductor does not expel the field B from any part of its bulk. Materials with such properties are called type-II superconductors and are quite significant for applications: indeed, magnets can be produced with large magnetic fields originated by large currents, that do not heat up because of thermal dissipation.

The behavior of superconductors in response to magnetic fields has one more crucial consequence. Suppose to have a hollow superconducting cylinder at $T > T_c$. When a magnetic field is applied and the temperature is lowered below T_c , the induction field B is expected to vanish inside the material, but not in the hollow part. Inside the hollow part, the magnetic flux persists even in the absence of the external magnetic field, though the current in the cylinder persists as well. To explain this observation, London assumed that the magnetic flux Φ_n be quantized according to the relation

$$\Phi_n = n\Phi_0. \quad (2.241)$$

That is, the magnetic flux is an integer n multiple of an elementary quantum $\Phi_0 = hc/e^*$, with e^* an effective charge, experimentally found to be consistent with the value $e^* = -2e$. Thus, electrons behave as if they were paired. In fact, other suited experiments lead to conclusion that electron pairs actually strongly overlap with each other.

Isotopic effect. Finally, studies of superconductivity in materials where atoms were substituted by isotopes, the critical field H_0 was found to depend on the mass M of the atoms according to $H_0 \propto 1/\sqrt{M}$. The same occurs for the critical temperature, that scales as $1/\sqrt{M}$.

In essence,

Concept

Zero-resistance behavior, transparency to radiation with frequency below a critical ω_c , exponential behavior of specific heat and thus for activation of thermal excitations, and perfect diamagnetism below T_c , seem to be compatible with the existence of an energy gap $\Delta(T)$ between the ground and the first excited single-particle states, that vanishes at and above the critical temperature. These experimental findings, especially the perfect diamagnetic behavior and the flux quantization, support the idea that the fermionic current-carriers are somehow paired.

Expression (2.239) may indeed explain why $\hbar\omega_c \simeq 4k_bT_c$ and suggests that an energy $2\Delta_0$ is needed to break a pair. However, the fact that these pairs are experi-

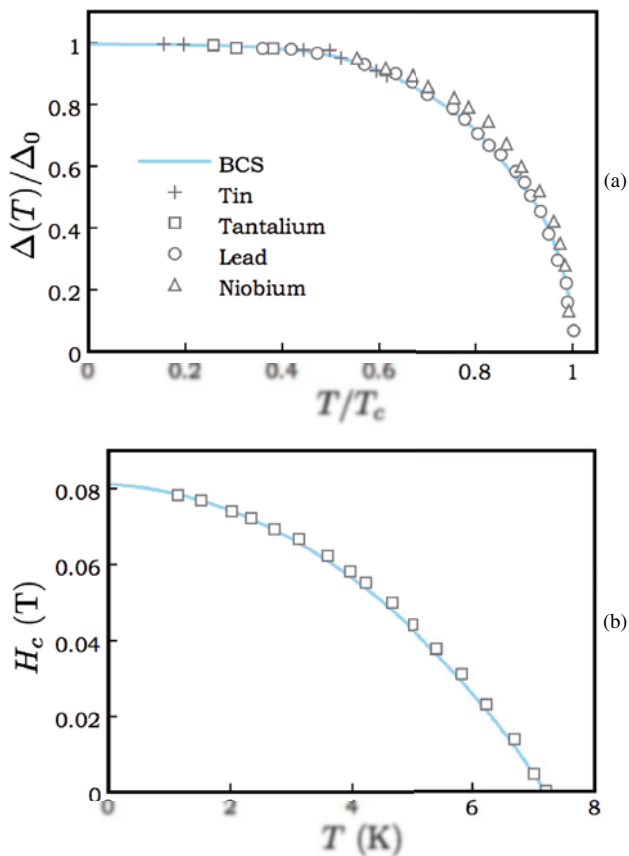


Fig. 2.47 Phenomenology of metallic superconductors. (a) Energy gap $\Delta(T)$ scaled to the $T = 0$ value Δ_0 vs. temperature T/T_c for a number of compounds as in the legend (symbols). Solid line: BCS result [38]. (b) Critical magnetic field vs. temperature T for Pb, having $T_c \simeq 7.2$ K and $H_0 \simeq 0.08$ T [39]

mentally inferred to be highly overlapped in real space, suggests that the gap energy does not coincide with the pair binding energy. In summary, a theory is needed to understand how the phenomenological energy gap is related to the microscopic pairing mechanism.

Also,

Concept

In regard to the specific case of metallic superconducting compounds, the manifestation of the isotope effect seems to be a fingerprint that ion vibrations play a significant role in mediating the pairing mechanism.

2.8.3 Phenomenology of superfluids and the two-fluid model

Before proceeding to assess the basic ingredients of the microscopic BCS theory for superconductivity, a phenomenological model is here illustrated, where at any given temperature part of the system is superfluid and the remaining part behaves as a normal fluid. The superfluid component with mass m_s (and if charged with charge e_s), density n_s and velocity \mathbf{v}_s , is the condensed quantum state of the Bose liquid and, as a single quantum state, carries no entropy. The normal component, with mass m_n (and if charged with charge e), density n_n and velocity \mathbf{v}_n , carries the whole thermal excitation of the fluid. The total density is $n = n_s + n_n$, the total current $\mathbf{J} = \mathbf{J}_s + \mathbf{J}_n = n_s \mathbf{v}_s + n_n \mathbf{v}_n$, the total entropy $nS = n_n S_n$ is only due to the normal component, with S and S_n the entropy per particle. At zero temperature the whole system is superfluid and above T_c the system is entirely normal. The model applies to both neutral superfluids and charged superconductors, in both cases irrespective of the boson or fermion nature of the composing particles and without need to specify the microscopic mechanism that originates the pairing. For Bose superfluids, the two-fluid model had the striking merit of successfully explaining the observation of density fluctuations different from ordinary sound waves: these correspond to propagation of local temperature fluctuations, leading to out-of-phase oscillations of the superfluid and normal-fluid components, named second sound. In ordinary (named first, in this context) sound, the normal and superfluid components execute in-phase oscillations instead.

In the following, the phenomenological model is discussed for superconductors. To see how it works, the key starting idea is to begin with looking at the difference in the density of free energies $f(\mathbf{r})$ between the superconducting and the normal state. This is given by

$$f(\mathbf{r}) = \alpha(T) |\psi(\mathbf{r})|^2 + \frac{1}{2} \beta(T) |\psi(\mathbf{r})|^4, \quad (2.242)$$

with $\alpha(T)$ and $\beta(T)$ temperature-dependent specific functions characteristic of the material and

$$|\psi(\mathbf{r})|^2 = \frac{n_s(\mathbf{r})}{n} \quad (2.243)$$

the fractional superfluid density n_s as compared to the total density n . It is assumed that the system be completely superfluid at $T = 0$ with $|\psi(\mathbf{r})|^2 = 1$, and be normal with $|\psi(\mathbf{r})|^2 = 0$ at $T = T_c$. Assuming that $f(\mathbf{r})$ is a slowly varying function of space, at the thermodynamical equilibrium it must occur that

$$\frac{\partial f}{\partial |\psi|^2} = 0 = \alpha(T) + \beta(T) |\psi|^2. \quad (2.244)$$

The stationary condition (2.244) provides the equilibrium density

$$|\psi_{eq}(T)|^2 = -\frac{\alpha(T)}{\beta(T)}, \quad (2.245)$$

so that $f_{eq}(T) \equiv f_S(T) - f_N(T) = -\alpha^2(T)/(2\beta(T))$. Since the change of free energy at the critical point corresponds to the density of magnetic energy when the field is expelled from the material, one also has that $f_{eq} = f_S(T) - f_N(T) = -H_c^2(T)/(8\pi)$. Therefore,

$$\frac{H_c^2(T)}{8\pi} = \frac{1}{2} \frac{\alpha^2(T)}{\beta(T)}. \quad (2.246)$$

A second equation comes from considering that

Concept

The superfluid motion corresponds to a non-viscous flow.

If \mathbf{v}_s is the velocity of the superfluid with density n_s , mass m_s , and charge e_s in the presence of an electric field one has the equation of motion $m_s \partial \mathbf{v}_s / \partial t = e_s \mathbf{E}$. Thus, for the current $\mathbf{J}_s = n_s e_s \mathbf{v}_s$ one has

$$\frac{\partial \mathbf{J}_s}{\partial t} = \frac{n_s e_s^2}{m_s} \mathbf{E}. \quad (2.247)$$

Using Maxwell equations, this turns into

$$\frac{\partial (\nabla \wedge \mathbf{J}_s)}{\partial t} = \frac{n_s e_s^2}{m_s} \nabla \wedge \mathbf{E} = -\frac{n_s e_s^2}{m_s c} \frac{\partial \mathbf{B}}{\partial t}, \quad (2.248)$$

or else

$$\frac{\partial}{\partial t} \left(\nabla \wedge \mathbf{J}_s + \frac{n_s e_s^2}{m_s c} \mathbf{B} \right) = 0. \quad (2.249)$$

The general solution of (2.249) is

$$\nabla \wedge \mathbf{J}_s + \frac{n_s e_s^2}{m_s c} \mathbf{B} = \mathbf{a} \quad (2.250)$$

with \mathbf{a} a constant quantity, independent of time. From now on, this is taken to be zero. Since $\mathbf{J}_s = (4\pi)^{-1} c \nabla \wedge \mathbf{B}$, (2.250) implies that

$$\nabla^2 \mathbf{B} = \frac{4\pi n_s e_s^2}{m_s c^2} \mathbf{B}. \quad (2.251)$$

In the simplest one-dimensional case, it is immediately found by inspection that $\mathbf{B}(\mathbf{r})$ behaves according to $\mathbf{B}(\mathbf{r}) = \mathbf{B}_0 e^{-r/\lambda_L}$ in terms of the quantity

$$\lambda_L(T) = \sqrt{\frac{m_s c^2}{4\pi n_s e_s^2}}. \quad (2.252)$$

The quantity λ_L is named London penetration depth and represents the length over which the induction field \mathbf{B} penetrates inside the material. Its measure infers infor-

mation on $n_s(T)$, as given for example in the Uemura plot, Fig. 2.45. Indeed one has

$$\frac{\lambda_L^2(0)}{\lambda_L^2(T)} = \frac{|\psi_{eq}(T)|^2}{|\psi_{eq}(0)|^2} = |\psi_{eq}(T)|^2 = -\frac{\alpha(T)}{\beta(T)} \quad (2.253)$$

Solution of (2.246) and (2.253) provides the functions $\alpha(T)$ and $\beta(T)$:

$$\alpha(T) = -\frac{H_c^2(T)}{4\pi} \frac{\lambda_L^2(T)}{\lambda_L^2(0)} \text{ and } \beta(T) = \frac{H_c^2(T)}{4\pi} \frac{\lambda_L^4(T)}{\lambda_L^4(0)}. \quad (2.254)$$

Notice that $\mathbf{B} = \nabla \wedge \mathbf{A}$ in terms of the vector potential \mathbf{A} , so that equation (2.250) can be cast in the form $\nabla \wedge (\mathbf{J}_s + \frac{n_s e_s^2}{m_s c} \mathbf{A}) \equiv \nabla \wedge \mathbf{J}'_s = 0$, simply stating that the superfluid is irrotational: the superfluid current $\mathbf{J}'_s = n_s e_s \mathbf{v}'_s$ and the superfluid velocity \mathbf{v}'_s have to be the gradient of a scalar quantity. From now on, the position $\mathbf{v}'_s = \hbar \nabla \phi / m_s$ is set, with ϕ representing a global phase of the superfluid wavefunction.

As to the quantization of flux, recalling the physical situation that originates (2.241), it is necessary to consider the integral

$$\int_C \mathbf{J}_s(\mathbf{r}) d\ell = (e_s n_s / m_s) \int_C [\hbar \nabla \phi - e_s \mathbf{A} / c] \cdot d\ell, \quad (2.255)$$

over a closed circuit C well inside the hollow cylinder. The vector potential \mathbf{A} has been conveniently introduced after recalling in (2.250) that $\mathbf{B} = \nabla \wedge \mathbf{A}$. The current inside the hollow is zero and thus the integral has to vanish. On the other hand, one has $\int_C \nabla \phi \cdot d\ell = 2\pi p$, stating that the phase change on a closed line can only be an integer multiple p of 2π . Considering that $\int_C \mathbf{A} \cdot d\ell = \int_S \mathbf{B} \cdot d\mathbf{S} = \Phi$ is by definition the flux of \mathbf{B} through the surface \mathbf{S} , one finally has

$$\Phi = p \frac{\hbar c}{e_s}, \quad (2.256)$$

with p an integer.

As shown within the BCS theory from microscopic arguments, consistency of the above equations with the experimental data requires that the density n_s , mass m_s , and charge e_s be substituted with effective values $n_s = n/2$, $m_s = 2m$, and $e_s = -2e$ respectively. Thus, the thermodynamical, electric and magnetic properties of a superconductor can be phenomenologically understood by viewing its superconducting electrons, or else charged fermions, as if they were paired, having a density dictated by the squared modulus of a macroscopic coherent wavefunction with a global phase, and an irrotational velocity given by the gradient of a global wavefunction phase.

Charged vs neutral superfluids. The Maxwell equations for a superconductor derived so far have a striking analogue for neutral superfluids, such as ^4He among bosonic-like systems or ^3He among fermionic ones. These are equations of motion for the total current density \mathbf{J} and for the superfluid velocity \mathbf{v}_s , named two-fluid equations [40, 41, 51]. The analogies are in the following put forward.

Like for Maxwell equations, also the form of these two-fluid equations of motion is dictated by general considerations about symmetries and conservation laws. As for a normal, i.e. non superfluid fluid, invariance under a Galileian transformation must hold, along with time-reversal symmetry, as well as conservation of linear and angular momentum. To these general considerations valid for normal fluids, the peculiar irrotational character of the superfluid field \mathbf{v}_s is to be added, at least below the threshold for formation of vortexes. The forces determining the matter and the energy flows are originated within the system by fluctuations in the density of particles and in the density of entropy: they work as conservative and restoring forces. The dissipative forces are related to spatial changes in the velocity of the normal fluid component \mathbf{v}_n and to spatial changes in the so called interdiffusion particle current $\mathbf{J}_r = n_s(\mathbf{v}_s - \mathbf{v}_n)$. In fact, $\mathbf{J}_r = n_s\mathbf{v}_s + n_n\mathbf{v}_n - n\mathbf{v}_n = n_s(\mathbf{v}_s - \mathbf{v}_n)$ is the current as seen in a reference frame moving with the normal fluid component. The restoring forces originate the dispersion, that is the wavevector dependence, of the superfluid excitation modes discussed at the beginning of this section and in Appendix 2.13. The dissipative forces determine the attenuation of these modes through viscosity coefficients. Four coefficients are bulk-like in nature and describe the mutual $\mathbf{J} - \mathbf{J}$, $\mathbf{J} - \mathbf{v}_s \approx \mathbf{v}_s - \mathbf{J}$, and $\mathbf{v}_s - \mathbf{v}_s$ interactions; one is shear-like in nature and involves only the normal component. In a normal -one component - fluid, only one bulk and one shear viscosity would occur. A more detailed demonstration of these ideas is provided in Chap. 6.

These concepts are quite general and can be adapted to both neutral and charged systems, once the appropriate definitions of density and currents are identified [41, 51]. Obviously, in neutral systems the charge e_s is substituted by the mass m_s , so that for example currents are generally written as $\mathbf{J} = nm\mathbf{v}$, with n the particle density. The definition and meaning of superfluid velocity, that is the irrotational part of the velocity field, remains unchanged. Identifying the analogue of the vector potential \mathbf{A} in Maxwell equations is slightly more elaborate. A vector potential essentially creates rotating currents: the question is which physical quantity would originate a stirring of the neutral fluid. To answer the question, consider the Hamiltonian term providing a rotation with angular velocity $\boldsymbol{\omega}$ below the threshold for vortex formation: this is $\int d\mathbf{r} \boldsymbol{\omega} \cdot \mathbf{L}$ in terms of angular momentum $\mathbf{L} = \mathbf{r} \wedge \mathbf{J}$. In a charged system, the minimal coupling form determining a rotational current is $\mathbf{J} \cdot \mathbf{A}$. Using the properties of vector and scalar products, the mapping $\mathbf{A} = \boldsymbol{\omega} \wedge \mathbf{r}$ is immediately suggested for the transverse vector potential.

Pushing the analogy farther, it can be also noticed that in charged systems the magnetic \mathbf{B} and electric \mathbf{E} fields are determined once a vector potential \mathbf{A} and a scalar potential, say V , are given. \mathbf{B} is essentially related to the curl of \mathbf{A} and \mathbf{E} to the gradient of V . Thus, one first notices that the angular velocity in the neutral system works as the magnetic field in the charged one, providing fluid stirring. Then, it is seen that the quantity $V + \hbar \partial \phi / \partial t$ is gauge invariant: writing the equations in the gauge in which the scalar potential vanishes corresponds to pass, by Galilean transformation, to a reference frame moving with $\mathbf{v}_s = \hbar \nabla \phi / m_s$. In such a frame, the quantity $\partial \mathbf{A} / \partial t + \nabla V / m_s$ is therefore gauge invariant. Using the analogy for \mathbf{A} and considering that only the normal component rotates, one has that the quantity

$\partial(\mathbf{v}_n - \mathbf{v}_s)/\partial t$ is gauge invariant. It is then seen that the Maxwell equation $\nabla \wedge (\mathbf{E} + c^{-1}\partial\mathbf{A}/\partial t) = 0$ or else $\mathbf{E} + c^{-1}\partial\mathbf{A}/\partial t = -\nabla V$ is just the condition for irrotational flow. Finally, after a few manipulations it is seen that the analogue of the electric field \mathbf{E} is $(n_s c)^{-1}\partial\mathbf{J}_r/\partial t$, and that the Maxwell equation for $\nabla \wedge \mathbf{B}$ expresses the relation of continuity linking particle and current densities,

With no need to know microscopic interactions or pairing mechanisms, the two-fluid model is a convincing phenomenological framework to establish the links between superconductivity and superfluidity in both charged and neutral, fermionic or bosonic, fluids. In fact, a microscopic derivation of the two-fluid model within a full quantum many-body treatment can be set, which is the subject of Time-Dependent Density Functional Theory for superfluids discussed in Chap. 6. There, the connection between the microscopic description of interactions and BEC or superfluidity is more clearly evident, as well as the difference between condensate and superfluid is more easily set:

Concept

The condensate density is related to the macroscopic occupation of the lowest energy state. The superfluid density in essence describes the response of the fluid to a twist, that is a transverse probe. Strong interactions affect the occupation of the lowest energy state and the condensate can be correspondingly depleted.

Ideally, a homogeneous superfluid at zero temperature is characterized to have a 100% superfluid density, but not necessarily a 100% condensate density.

In the next section the focus is reverted back to the microscopic nature of the superconducting ground and excited states of fermionic systems, by investigating the case of metallic superconducting compounds.

2.8.4 BCS theory

The phenomenology illustrated so far suggests that the manifestation of superconductivity be originated by a microscopic pairing mechanism in the presence of an attractive interaction. The Bardeen-Cooper-Schrieffer theory sets the theoretical framework and tools to handle the complex problem. While most of the following treatment is general, actual calculations require some modeling. Whenever introduced, specific modeling is referred to the metallic superconducting compounds, where the microscopic pairing mechanism is due to phonons, the corresponding energies being on the Debye scale $\hbar\omega_D$, as introduced in Chapter 3.

In order to proceed gradually in these complex matters, the problem of two electrons in the presence of the attractive interactions due to phonons is first illustrated, as derived by Cooper. The formation of Cooper pairs is then discussed for the many-electron system: to this aim, the pedagogical approach [4] is here followed to derive the equations at an easier level.

2.8.4.1 Cooper pairs

Consider two free electrons with opposite spins and at opposite sides of the Fermi surface, thus described by plane waves with wavevectors \mathbf{k} and $-\mathbf{k}$, in the presence of an attractive interaction. The Hamiltonian $H = H_0 + U$ is composed by the kinetic term H_0 for the two electrons and this effectively attractive interaction term U between them. Though the specific form of this attractive interaction changes according to system characteristics, at this level it is sufficient to state that the interaction be attractive and effective for energies smaller than the typical interaction energy range E_a which originates the attraction. Therefore, it can be modeled to be $U(\varepsilon) = -V$ for energies ε smaller than E_a and zero for $\varepsilon > E_a$. In the specific case of metallic superconducting compounds, the microscopic nature of such an attraction is due to the interaction between the two electrons and one acoustical phonon, so that $E_a = \hbar\omega_D$ is the typical Debye energy. The detailed derivation of the microscopic interaction is executed in Sec. 3.11 of the next Chapter.

The problem is simplified if the pair motion is separated into a center-of-mass and a relative-coordinate contribution. The center of mass is not subjected to any force and therefore moves with constant momentum \mathbf{Q} that, in this case, is $\mathbf{Q} = 0$. The relative motion is more elaborated, since this is precisely determined by the internal force, here given by the effective attraction. In the absence of interactions, the wavefunction of the pair would be the plane wave $\phi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}$ as a function of the relative coordinate $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, and its energy would be twice $\varepsilon_{\mathbf{k}} = \hbar^2 k^2 / (2m)$. In the presence of the attractive interaction, the pair wavefunction can be built from a combination of the above plane waves, provided that the corresponding wavevectors and energies be larger than ε_f to account for the Pauli exclusion principle and the fact that states below ε_f are all occupied. Thus, the pair wavefunction in the relative coordinates can be cast in the form

$$\Psi(\mathbf{r}) = \sum_{\mathbf{q}} a(\mathbf{q}) \phi_{\mathbf{q}}(\mathbf{r}). \quad (2.257)$$

with $|\mathbf{q}| \geq k_f$ as already noticed. The Schrödinger equation $H\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$ for $\Psi(\mathbf{r})$ leads to the equation for the coefficients $a(\mathbf{q})$:

$$(2\varepsilon_{\mathbf{k}} - E)a(\mathbf{q}) + \sum_{\mathbf{q}} a(\mathbf{q}) \langle \phi_{\mathbf{k}} | U | \phi_{\mathbf{q}} \rangle = 0 \quad (2.258)$$

The matrix element $\langle \phi_{\mathbf{k}} | U | \phi_{\mathbf{q}} \rangle$ can be calculated after inserting the model expression for U and integrating over \mathbf{q} . The resulting expression is especially helpful when the integral is cast over the density of states, that is:

$$(2\varepsilon - E)a(\varepsilon) - V \int_{\varepsilon_f}^{\varepsilon_f + \hbar\omega_D} d\varepsilon' g(\varepsilon') a(\varepsilon') = 0, \quad (2.259)$$

where $-V$ is the attractive potential in the energy range $\varepsilon_f - \hbar\omega_D < \varepsilon < \varepsilon_f + \hbar\omega_D$. Indeed, the density of states $g(\varepsilon)$ does not significantly change in the range of en-

ergies $\hbar\omega_D \ll \varepsilon_f$, so that one may approximate $g(\varepsilon) \simeq g(\varepsilon_f)$ to be the (constant) value that it has at the Fermi energy and get out of the integral. An expression for the eigenvalue E is yet to be found, that represents the energy of the pair. To this aim, one notices that the coefficients $a(\varepsilon)$ scale as $a(\varepsilon) \approx 1/(2\varepsilon - E)$, provided that

$$E = 2\varepsilon_f + 2 \frac{\hbar\omega_D}{1 - e^{2/[g(\varepsilon_f)V]}}. \quad (2.260)$$

Under the conditions here considered, $e^{2/[g(\varepsilon_f)V]} \gg 1$ and thus

$$E = 2\varepsilon_f - 2\hbar\omega_D e^{-2/[g(\varepsilon_f)V]}. \quad (2.261)$$

Expression (2.261) comes to the relevant conclusion that the total energy E of the pair in the presence of the attraction $-V$ is lower than the energy of the non-interacting pair. Therefore, this suggests that the process of switching on a whatever weak attraction between the two electrons, changes the nature of the ground state. In this two particle problem, the quantity $2\hbar\omega_D e^{-2/[g(\varepsilon_f)V]} \equiv \Delta_0$ is the energy needed to break the pair, that is its binding energy. In the many-electron problem, this quantity evolves into the energy gap that opens up in the single-particle spectrum. The derivation of this result is the aim of the next section.

2.8.4.2 The many- particles BCS state

BCS theory stems from the consideration that a pairing mechanism exists due to some attractive interaction potential, so that the ground state of a superconductor be built up from a Slater determinant of pair wavefunctions, with reference to Appendix 2.14. In the conventional, metallic superconductors which the original BCS theory was referred to, isotope effect and other experiments indicate that the pairing mechanism is due to the interaction among electrons and lattice vibrations: as discussed in Sec. 3.11.2 of Chap. 3, this leads to an effective electron-electron interaction mediated by these lattice vibrations and that is independent of the electrons spins. The BCS theory though, is a sufficiently general framework to be adapted to other attractive interactions originated by different microscopic mechanisms. In the following, the general situation is considered, unless otherwise specified.

Reverting back to the BCS ground-state wavefunction, this is assumed to be of the type

$$|\Psi\rangle = \prod_{\mathbf{k}} (u_{\mathbf{k}} + v_{\mathbf{k}} c_{\mathbf{k}\uparrow}^{\dagger} c_{-\mathbf{k}\downarrow}^{\dagger}) |0\rangle, \quad (2.262)$$

The vacuum $|0\rangle$ is in essence filled by creating pairs of electrons where one electron has momentum \mathbf{k} and spin up and the other $-\mathbf{k}$ and spin down. Momenta are opposite to ensure the translational invariance of the lowest energy state, since there is no kinetic energy associated with a drift of the pairs. The state (2.262) can be viewed as a generalization in \mathbf{k} -space to the many body problem of the single Cooper pair treated in the previous section. Spins are to be opposite for the pair to be in a singlet state and ensure diamagnetic behavior. Triplet paired states are possible, as in ^3He ,

and more complicated equations arise in that case. The functions $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ are to be determined, along with the normalization condition $|u_{\mathbf{k}}|^2 + |v_{\mathbf{k}}|^2 = 1$.

As to the Hamiltonian, this is taken in the general form:

$$H = \sum_{\mathbf{k},s} \xi_{\mathbf{k}} c_{\mathbf{k}s}^\dagger c_{\mathbf{k}s} + \sum_{\mathbf{k},\mathbf{k}'} V_{\mathbf{k}'\mathbf{k}} c_{\mathbf{k}'\uparrow}^\dagger c_{-\mathbf{k}'\downarrow}^\dagger c_{-\mathbf{k}\downarrow} c_{\mathbf{k}\uparrow}, \quad (2.263)$$

where s labels the two spin orientations, the sums over \mathbf{k} and \mathbf{k}' are done in the energy range $\varepsilon_f - \Delta E < \varepsilon < \varepsilon_f + \Delta E$, and $2\Delta E$ is the range of the attractive interaction. For metallic superconductors, this would be in the range of typical lattice vibrational energies. In essence, (2.263) states that two electrons with opposite momenta \mathbf{k} and $-\mathbf{k}$ and opposite spins are destroyed, and two created with momenta \mathbf{k}' and $-\mathbf{k}'$ without any spin flip. The missing momentum $\mathbf{k}' - \mathbf{k}$ to ensure momentum conservation is provided by the excitation that mediates the pairing (a lattice vibration in metallic superconductors). Notice that conservation of particle number is instead ensured by the introduction of the chemical potential through the appearance of the $-\mu \sum_{\mathbf{k},s} c_{\mathbf{k},s}^\dagger c_{\mathbf{k},s}$ piece in the kinetic energy. The sum over \mathbf{k} has to be performed in a way that $\xi_{\mathbf{k}} = \varepsilon_{\mathbf{k}} - \mu$ if $|\mathbf{k}| > k_f$ and $\xi_{\mathbf{k}} = -(\varepsilon_{\mathbf{k}} - \mu)$ if $|\mathbf{k}| < k_f$.

The calculations to find eigenvectors and eigenvalues of Hamiltonian (2.263) can be carried out starting from (2.262) and (2.263), determining the coefficients $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ by means of a variational method. To simplify the equations, the following model for the attractive potential is once more used from now on:

$$\begin{aligned} V_{\mathbf{k}\mathbf{k}'} &= -V < 0 \text{ if } |\varepsilon_{\mathbf{k}} - \varepsilon_f|, |\varepsilon_{\mathbf{k}'} - \varepsilon_f| < \Delta E \\ V_{\mathbf{k}\mathbf{k}'} &= 0 \text{ otherwise.} \end{aligned} \quad (2.264)$$

That is, an attractive interaction exists between electrons with energies $\varepsilon_{\mathbf{k}}$, $\varepsilon_{\mathbf{k}'}$ lying in a tiny shell of width ΔE around the Fermi energy ε_f .

An approximate diagonalization of the BCS Hamiltonian can be obtained after applying the Bogolubov approximation. This amounts to define operators that are linear combinations of the c 's:

$$\begin{aligned} d_{\mathbf{k}\uparrow}^\dagger &= u_{\mathbf{k}} c_{\mathbf{k}\uparrow}^\dagger - v_{\mathbf{k}} c_{-\mathbf{k}\downarrow} \\ d_{-\mathbf{k}\downarrow} &= u_{\mathbf{k}} c_{-\mathbf{k}\downarrow} + v_{\mathbf{k}} c_{\mathbf{k}\uparrow}^\dagger, \end{aligned} \quad (2.265)$$

creating or destroying two types of excitations in the new ground state, that are described by $d_{\mathbf{k}\uparrow}^\dagger|0\rangle$ and $d_{-\mathbf{k}\downarrow}^\dagger|0\rangle$. The new ground state is chosen to not contain such excitations, so that $d_{\mathbf{k}\uparrow}|0\rangle = 0$. Notice that the functions $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ have been chosen to be real, without loss of generality. Now, one has to express the BCS Hamiltonian in terms of the d 's and impose that the Hamiltonian be diagonal in such operators. When the c operators are derived from (2.265) and substituted in the BCS Hamiltonian, the destruction d operators can be conveniently moved to the right with the help of the Fermi anticommutation relations $\{d_{\mathbf{k}s}, d_{\mathbf{k}'s'}^\dagger\} = \delta_{\mathbf{k}\mathbf{k}'} \delta_{ss'}$. Using the properties of the d operators when acting on $|0\rangle$, the following expression is at the end left out:

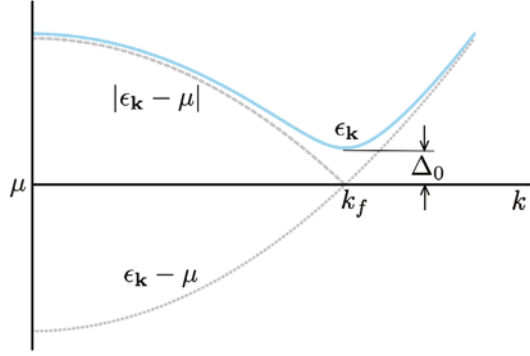


Fig. 2.48 The single-particle excitation energy E_k for a superconductor, showing the gap at $k = k_f$ and the free-particle limit while $k \rightarrow \infty$

$$H = 2 \sum_{\mathbf{k}} \xi_{\mathbf{k}} v_{\mathbf{k}}^2 - V \sum_{\mathbf{k}\mathbf{k}'} u_{\mathbf{k}} v_{\mathbf{k}} u_{\mathbf{k}'} v_{\mathbf{k}'} + \sum_{\mathbf{k}} [2\xi_{\mathbf{k}} u_{\mathbf{k}} v_{\mathbf{k}} - V(u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2) \sum_{\mathbf{k}'} u_{\mathbf{k}'} v_{\mathbf{k}'}] d_{\mathbf{k}\uparrow}^\dagger d_{-\mathbf{k}\downarrow}^\dagger. \quad (2.266)$$

The last term is to vanish, so that $2\xi_{\mathbf{k}} u_{\mathbf{k}} v_{\mathbf{k}} = V(u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2) \sum_{\mathbf{k}'} u_{\mathbf{k}'} v_{\mathbf{k}'}$. From the normalization condition $u_{\mathbf{k}}^2 + v_{\mathbf{k}}^2 = 1$ and using the definition $\Delta_0 \equiv V \sum_{\mathbf{k}} u_{\mathbf{k}} v_{\mathbf{k}}$, one finds

$$u_{\mathbf{k}}^2 = \frac{1}{2} \left(1 + \frac{\xi_{\mathbf{k}}}{E_{\mathbf{k}}} \right) \quad \text{and} \quad v_{\mathbf{k}}^2 = \frac{1}{2} \left(1 - \frac{\xi_{\mathbf{k}}}{E_{\mathbf{k}}} \right) \quad (2.267)$$

with

$$E_{\mathbf{k}} = \sqrt{(\epsilon_{\mathbf{k}} - \mu)^2 + \Delta_0^2}. \quad (2.268)$$

The latter can be identified with the single-particle excitation energy and is displayed in Fig. 2.48. This is the energy needed to destroy one pair and create one excitation, as shown in the following. Since the ground-state energy is $E_{gs} = 2 \sum_{\mathbf{k}} \xi_{\mathbf{k}} v_{\mathbf{k}}^2 - V \sum_{\mathbf{k}\mathbf{k}'} u_{\mathbf{k}} v_{\mathbf{k}} u_{\mathbf{k}'} v_{\mathbf{k}'} = \sum_{\mathbf{k}} (2\xi_{\mathbf{k}} v_{\mathbf{k}}^2 - \Delta_0 u_{\mathbf{k}} v_{\mathbf{k}})$, the energy required to destroy one pair is $2\xi_{\mathbf{k}} v_{\mathbf{k}}^2 - 2\Delta_0 u_{\mathbf{k}} v_{\mathbf{k}}$. Therefore, the energy required to get one particle excited is $\xi_{\mathbf{k}} - 2(\xi_{\mathbf{k}} v_{\mathbf{k}}^2 - \Delta_0 u_{\mathbf{k}} v_{\mathbf{k}}) = E_{\mathbf{k}}$, where the last equality results from using (2.267)-(2.268). As it is evident, the excitation energy is characterized by a finite gap even at $k = 0$, meaning that a finite energy is needed to create an excitation.

Once the definition $\Delta_{\mathbf{k}} = u_{\mathbf{k}} v_{\mathbf{k}}$ is used and substituted into $2\xi_{\mathbf{k}} u_{\mathbf{k}} v_{\mathbf{k}} = V(u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2) \sum_{\mathbf{k}'} u_{\mathbf{k}'} v_{\mathbf{k}'}$, the following coupled equations result for the gap function and for the total number N of particles (pairs):

$$\begin{aligned} \Delta_{\mathbf{k}} &= - \sum_{\mathbf{k}'} V_{\mathbf{k}\mathbf{k}'} \frac{\Delta_{\mathbf{k}'}}{2E_{\mathbf{k}'}} \\ N &= \sum_{\mathbf{k}'} 2v_{\mathbf{k}}^2. \end{aligned} \quad (2.269)$$

The two equations above are to be solved self-consistently. To this aim, numerical methods are conveniently used: the algorithm is started with a guess for $\Delta_{\mathbf{k}'}$ on a grid of wavevectors, and the first equation (2.269) is iterated while requiring that

the second be satisfied, unless the desired convergence is reached. In performing the sums, the caveat is to be adopted of limiting them to values of k such that the condition that $\xi_{\mathbf{k}}, \xi_{\mathbf{k}'} < \Delta E$ is fulfilled.

In the following, the specific case of metallic superconductors is treated, where ΔE is in the range of the energies of the vibrational lattice states. Microscopic calculations of the electron-vibrational states interactions lead to effective attractive expressions that can be modeled according to (2.264): in a pictorial view, a local lattice deformation induced by the presence of one electron is able to attract a second electron passing nearby. In the specific calculation thus, $\Delta E = \hbar\omega_D$ is set, with ω_D the so-called Debye frequency typical of vibrational states discussed in Chap. 3. Using the simplified expression for $V_{\mathbf{k}\mathbf{k}'}$, one obtains $1 = (V/2)\sum_{\mathbf{k}}(1/E_{\mathbf{k}})$ for the gap equation. In the limit $\Delta_0 \ll \hbar\omega_D \ll \varepsilon_f$, this yields

$$\Delta_0 \simeq 2\hbar\omega_D e^{-2/[g(\varepsilon_f)V]}, \quad (2.270)$$

in terms of the density of states $g(\varepsilon_f)$ at the Fermi level.

The condensation energy instead, that is the difference between the normal and superfluid ground-state energies is $E_N - E_S = g(\varepsilon_f)\Delta_0^2/2$, that is the energy per pair Δ_0 times the number of pairs $g(\varepsilon_f)\Delta_0/2$. Setting $E_N - E_S = H_c(T=0)^2/(8\pi)$ yields the relationship $H_0 \equiv H_c(T=0) = 2\sqrt{\pi g(\varepsilon_f)\Delta_0}$ between the zero temperature critical field to the gap energy. All these findings lead to identify the gap Δ as the order parameter for the phase transition from the normal to superconducting state. This is indeed the case. Along these lines, the pairing gap Δ_0 plays the same role as the macroscopic average Φ does in Bose-Einstein condensation. The difference is that $\Delta_0 = \langle \sum_{\mathbf{k}} c_{-\mathbf{k}\downarrow} c_{\mathbf{k}\uparrow} \rangle$ is calculated from two fermion operators arranged so that the spin and momentum in pairing overall sum up to zero, while $\Phi = \langle d_0 \rangle$ is calculated from one single boson-like operator with $\mathbf{k} = 0$, so that the condensate wavefunction corresponds to zero spin and zero momentum.

Finally, at finite temperature one has that $u_{\mathbf{k}}v_{\mathbf{k}}$ is to be modified to account for the occupation factors of the excitations with momentum \mathbf{k} and spin up, and with momentum $-\mathbf{k}$ and spin down: the first of (2.269) is replaced by

$$\Delta_{\mathbf{k}} = - \sum_{\mathbf{k}'} \frac{V_{\mathbf{k},\mathbf{k}'}\Delta_{\mathbf{k}'}}{2E_{\mathbf{k}'}} (1 - f_{\mathbf{k}'\uparrow} - f_{\mathbf{k}'\downarrow}). \quad (2.271)$$

The temperature-dependent gap $\Delta(T)$ is now given by

$$\Delta(T) = V \sum_{\mathbf{k}} \frac{V_{\mathbf{k}\mathbf{k}'}\Delta_{\mathbf{k}'}}{2E_{\mathbf{k}'}} \tanh\left(\frac{E_{\mathbf{k}'}}{2k_b T}\right). \quad (2.272)$$

In the case of the simplified attractive interaction and in the usual regime of small Δ_0 , one finds $\Delta(T) \approx \Delta_0 \sqrt{1 - (T/T_c)}$ close to T_c , with a discontinuity of the temperature derivative at T_c : this is the property of a second-order phase transition, namely with no latent heat involved. In addition, it can be found that $k_b T_c = 1.14\hbar\omega_D e^{-2/[g(\varepsilon_f)V]}$. This latter result accounts for two experimental find-

ings: the isotope effect on the critical temperature through the dependence of T_c on the energy $\hbar\omega_D$ of the vibrational mode and, after use of the exponential expression for Δ_0 , the relation between Δ_0 and $k_b T_c$, $2\Delta_0/(k_b T_c) \simeq 3.5$. Thus

Concept

The dependence of T_c on the isotopic species and on the vibrational energy supports the idea that the pairing mechanism in conventional superconductors is originated by lattice vibrations .

High-temperature superconductors. As it is seen, BCS theory with the model attractive interaction mediated by vibrational crystal modes accounts surprisingly well for all the experimental findings. This is not the case when treating high- T_c superconductors, for which a few basic phenomenological differences are to be taken into account [42]. The most relevant one is that the spatial extent of the Cooper pairs ξ dictating the so-called coherence length, is much smaller than in normal superconductors. In fact, while in the latter $\xi/r_0 \gg 1$ on the scale of the interparticle distance, $\xi \approx r_0$ in high- T_c s. Eagles [43] and Leggett [44] have shown on general grounds that the parameter ξ/r_0 drives a continuous evolution that can be established from the BCS ground state of fermion pairs to a BEC of point-like bosons composed of pairs of fermions. High- T_c superconductors therefore appear to be somewhere in this crossover from BCS to BEC, so that the BCS theory in its original specific formulation for conventional superconductors is not sufficient. Besides the size of the ξ/r_0 parameter in the superconducting state, high- T_c s are peculiar in their non-superconducting phase. First, their average carrier density is of the order of 10^{20} cm^{-3} as in heavily doped semiconductors. Second, charge carriers are typically holes resulting from the oxygen doping. In addition, their phase diagram includes at least antiferromagnetic ordering and other ground states without gaps or with gaps in the lowest energy excitation spectrum, different from the superconducting gap. Also, the dependence of resistivity on temperature reminds behaviors typical of strongly doped semiconductors and of systems with reduced dimensionality (mainly 2D). Finally, the spin symmetry of the pair state is that of a d -wave instead than being a singlet. Therefore, more microscopic pairing mechanisms could be expected to be involved, their typical interaction strength being larger than in conventional superconductors and more localized in real space to account for the shorter pair size. In spite of these complications, the general BCS framework along with the BCS-BEC crossover may at least model most of the experimental findings.

Quick Questions

Q17. Large Hadron Collider (LHC) at CERN is the worlds largest and most powerful particle accelerator. It is a 27-km ring of so powerful magnets that they are able to boost the energy of the particles along the way at the gigantic speeds and energies needed to probe for example the Higgs boson and to investi-

gate other fascinating problems, such as the asymmetry between the amounts of matter and antimatter in the Universe, or understand the nature of dark energy and dark matter, the matter we can detect forming just the 4% of Universe. Realizing so powerful magnets has been a major issue. Considering what you have understood from superconductivity, discuss which would be a key-idea to make them.

Answer. Strong magnets can be certainly manufactured by electromagnets with very large winding currents. However, the larger the current, the larger is the power dissipated. So that in practice, normal electromagnets may work with upper limited currents and thus produce magnetic fields with limited intensity. This would be not case if superconducting magnets are used instead: this is indeed the concept used in building LHC, though the actual realization is technically challenging.

Q18. BCS theory or appropriate modifications evolving toward BEC-like

ground states work independently of the microscopic pairing mechanism, the crucial issue being the existence of a pairing mechanism. Based on high- T_c phenomenology, list at least two different pairing mechanisms.

Answer. In their normal state, high- T_c superconductors are very polar insulators and are characterized by antiferromagnetic effects. One could therefore guess that one possible pairing mechanism involve optical vibrational modes of the lattice (as discussed in Chap. 3), or else antiferromagnetic magnon modes. Proposed microscopic theories include in fact these possibilities among others.

2.9 Experimental methods to determine electronic spectra in crystals and nanostructures

Different experimental methods can be exploited to probe electronic excitations. In essence, an electronic excitation has to be first created in the system by an incoming probe, and the system response is next determined by investigating the resulting process outcome. Along the reasoning outlined in Chap. 1 when discussing the experimental methods to investigate the crystal structure,

Concept

Incoming probes can be either photons or massive particles with energies above the 0.1 to 10 eV range of the electronic excitations under study, and wavelengths comparable to interparticle distances.

Here therefore, either photons in the range from ultraviolet to X-radiation, or particles with de Broglie wavelengths of the order of $\lambda \approx \text{\AA}$ are suited. Similarly,

the response is in terms of changes in energy and wavevector of outgoing either photons or massive particles. In an additional classification, different methods might involve either absorption, reflection, or emission processes, or again either elastic or inelastic scattering.

Concept

The massive particles may in turn be either charged or neutral. Electrons and neutrons are more often used. On general grounds, charged particles experience strong, long-ranged Coulomb interactions while entering the material and are therefore able to penetrate only for a few layers: scattering processes based on electrons thus work as surface-sensitive probes. Neutrons are instead uncharged and the strong-force interaction which they act with is very short-ranged: scattering processes based on neutrons thus work as bulk probes.

Both electrons and neutrons do have a $1/2$ spin and a magnetic dipole moment, so that they can undergo magnetic dipole-dipole interactions. One more difference is essential here. Electrons with wavelengths $\lambda \approx \text{\AA}$ have energies of the order of $E_e \approx h^2/(2m_e\lambda^2) \approx 150.8/(\lambda[\text{\AA}])^2 \text{eV} \approx$ several tens of eV. Neutrons instead would have $E_n \approx h^2/(2m_n\lambda^2) \approx 0.082/(\lambda[\text{\AA}])^2 \text{eV} \approx$ several tens of meV, that is comparable thermal energies $k_b T$ at room temperature, $\approx 25 \text{ meV}$. They are in fact called thermal neutrons. Therefore,

Concept

Electron scattering is suited to probe high-energy excitations such as collective electronic excitations and interband transitions, whereas thermal neutrons are most suited to investigate low-energy excitations such as the vibrational modes of the crystal treated in Chap. 3 or the spin-wave, magnon modes in ferro and antiferromagnets treated in this Chapter.

Two more crucial issues about the methods under discussion are the sample and the escape depth. The former is related to the emission of photons and is dictated to the inverse light-absorption coefficient, that at these frequencies is of the order of 200 \AA [45]. The latter is related to the emission of electrons, which depends on their energy according to a curve that is almost universal for most solids: it can be as small as a few \AA in the range between 50 and 100 eV, becoming larger for lower and higher energies, though not larger than $\approx 40 \text{ \AA}$ at 1 keV [45]. Along these lines, photons and neutrons are bulk probes, whereas electron probes significantly suffer surface effects.

Finally, changes in electron or hole velocities due to their mass or charge characteristics or in their spin, can be determined by zero-frequency electric or magnetic fields, leading to transport behavior. Details on the specific mechanisms are postponed to Chaps. 4 and 5, where the theoretical tools are built up that can deepen the understanding of the experimental working principles. Here just a sketch of the different possibilities is illustrated. The interested student is also referred to [45].

2.9.1 Transport processes

In methods based on transport processes, the concept is that changes of electric or thermal-energy currents set in after the application of suited external electric, magnetic, or temperature-gradient fields coupling to the particles mass, charge or spin. The system capacity to respond to such external drives embodies information on the characteristics of the involved current carriers. A measurement of such characteristics might provide insights on selected features of the electronic bands, as specified in Chap. 4.

2.9.2 Electron Energy Loss Spectroscopy

A different scheme is based on the concept that changes occur in the energies of incoming charged particles, after inelastic scattering with crystal excitations. This is for example the case of Electron Energy Loss Spectroscopy (EELS). In EELS, an electron beam enters the sample with a known and narrowly defined kinetic energy up to tens of keV, where it creates crystal excitations. The electron beam is scattered in part via elastic processes without energy loss, and in part via inelastic processes. A typical EELS sketched spectrum along with an example spectrum from three different carbon-based compounds is displayed in [Fig. 2.49](#). In inelastic processes against crystal excitations, energy can be lost by the electrons of the beam. A measure of the amount of lost energy provides information on the energy of the excitation. Inelastic interactions include collective electronic modes with typical energies in the range of 10-20 eV, as detailed in Chaps. 3 and 5. At higher energies, EELS provides information on valence and core electron states, bonding with different elements. In all cases, EELS is especially suited instead than light probes, whenever longitudinal electronic excitations are to be investigated: probes based on photons originate transverse excitations, since the electric field of the wave is perpendicular to the propagation direction.

2.9.3 Photo-Emission Spectroscopy

Photo-Emission spectroscopy (PES) uses the concept of photoelectric effect. Photons impinging on the sample are absorbed and excite an electron from a state below the Fermi level. Part of the photon energy is used to unbind the electron, and the remaining part as kinetic energy to fly away: once the energy of the incoming photon is given, electrons with smaller kinetic energy and thus larger binding energy are more likely ejected. These can be either ultraviolet as in the so-called UPS, or X-ray photons as in XPS. A measure of the amount of kinetic energy of the excited (photo)electron provides information about the binding energy of the state from which the electron has been originated and eventually probes the density of states

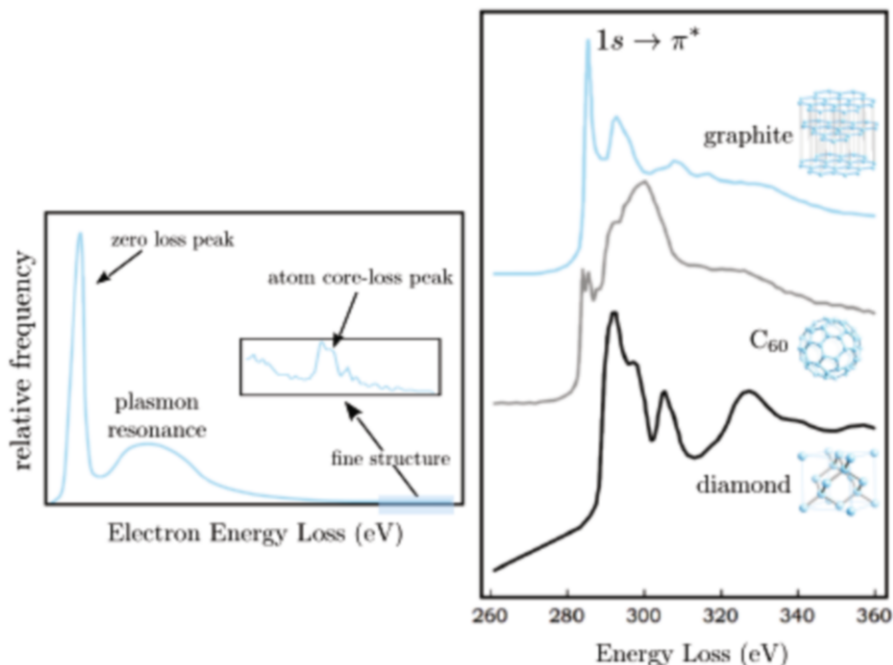


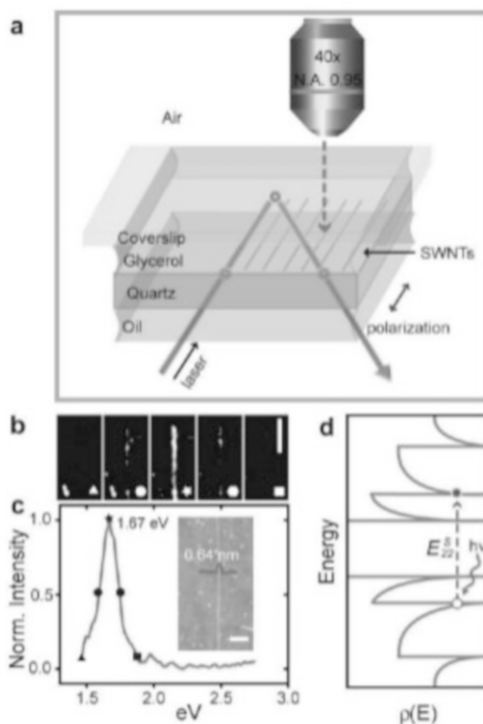
Fig. 2.49 Electron Energy Loss Spectroscopy. Left panel: schematic representation of EELS spectrum. From lower to higher energies, the narrow peak corresponds to elastic scattering, the broad shoulder to collective electronic excitations, such as the plasmons described in Chap. 4, the finer structures atom core-electrons scattering processes. Right panel: EELS spectrum for three carbon-based compounds. From top to bottom: graphite, C₆₀, and diamond. The three share the edge at 284 eV that is the fingerprint of carbon atom. The finer peak structures are different according to different bonding and local electronic state: in graphite, the peak corresponds to the transition between $1s$ occupied state and π antibonding empty state; the peak is absent in diamond, where no π electron exists [46]

below the Fermi level. UPS provides information on outer electron states and XPS on inner deep states.

PES is especially sensitive to surface effects: the excited electrons are indeed characterized by mean free paths of the order of a few nanometers and they need to escape the surface in order to be detected. In Inverse PES, electrons from the beam execute transitions to unoccupied states above the Fermi level after emitting a photon. The energy difference between the original beam electron and the final emitted photon provides information on the energy of unoccupied states. For this reason, a combinations of PES and IPES can be used to determine the band gap of a material. In Angle-Resolved Photoemission Spectroscopy (ARPES), the wavevector distribution of the outgoing photoelectrons is investigated besides the energy, providing information on the \mathbf{k} -dependence of the excitation energies.

Fig. 2.50 displays a schematic representation of a PES process along with the actual band structure of copper, as determined from ARUPS.

Fig. 2.51 Rayleigh imaging to measure optical scattering from individual single-walled carbon nanotubes (SWNT) on a solid substrate with high spatial and spectral resolution. (a) Schematic of the optical setup: the incident laser light impinges at an angle, travels through a set of index-matched media, and is reflected completely at the top interface having air as a lower refractive index medium. (b) Spatial Rayleigh spectrum snapshots at different frequencies. (c) Rayleigh scattering spectrum as a function of excitation energy: different symbols correspond to those labeling the different snapshots in (b). (d) Density of states $\rho(E)$ and van Hove singularities [49]



In this setting, the number of the absorbed photons is measured as a function of the incoming photon energy. The corresponding spectrum is in fact seen to be a smooth function of energy. On top of the smooth background spectrum, sharper edges can be superimposed corresponding instead to the ejection of core electrons. Evidently, these edges correspond to energies that are characteristic of each element: since the core electron is excited above the Fermi level, XAS probes the density of unoccupied states. The edge usually includes a jump preceded and followed by a series of oscillations typically depending on the number of neighbor atoms and related interactions. Though the obtained information is similar to that from EELS, XAS is not suited to investigate small sample areas since incident X-rays are hard to be focused. A schematic representation of the X-ray scattering processes is reported in Fig. 2.52.

2.9.6 Luminescence and X-ray Emission Spectroscopy

The inverse process of XAS is the X-ray Emission Spectroscopy (XES). This uses the concept of luminescence, occurring after that incoming X-rays have removed core electrons from the sample by photoelectric effect: an electron in an occupied

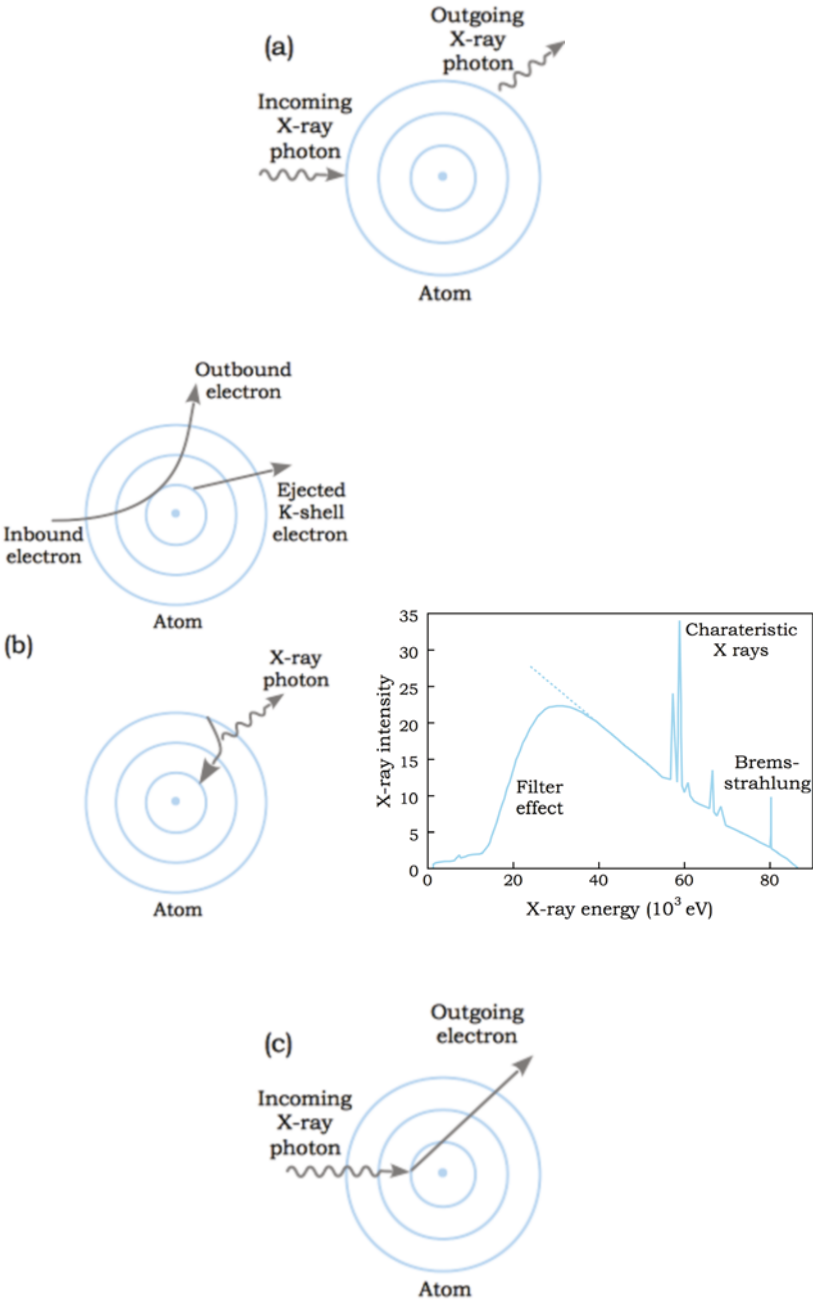


Fig. 2.52 Schematic representation of scattering processes from X rays, as described in the text. (a) Rayleigh scattering. (b) Fluorescence (left) and typical spectrum of characteristic fluorescence peaks on top of the background Bremsstrahlung radiation (right). (c) Photoelectric effect [1]

valence band can thus move down and occupy the just freed lower core level, emitting an X-ray. The density of states of conduction or valence bands, where the electrons move down from, can therefore be investigated by this process.

Luminescence is indeed the emission of light from a (cold) body after a quantum excitation, in contrast to incandescence that is generation of light from a (hot) body after a thermal excitation. Since the positions of the heavier nuclei are basically unchanged during the time of an electronic transition, the excitation energy of the system essentially accounts for the energy of the electronic excitation: this is the content of the Franck-Condon Principle.

In luminescence, the de-excitation of the sample may occur within tens of nanoseconds or within much longer times, even hours. The former process is referred as fluorescence and the latter as phosphorescence. In phosphorescence, energy is stored in metastable states and released via slower, typically thermally activated, processes. In a third type of luminescence, de-excitation occurs after chemical reactions: this is the chemiluminescence.

The energies of the emitted photons in the fluorescence process underlying XES, are sharply defined: they are indeed characteristic of the energy difference between the higher state which the electron moves down from, and the lower emptied state, where it moves to. Fig. 2.53 displays after [50] the photoluminescence spectrum of a single walled carbon nanotube.

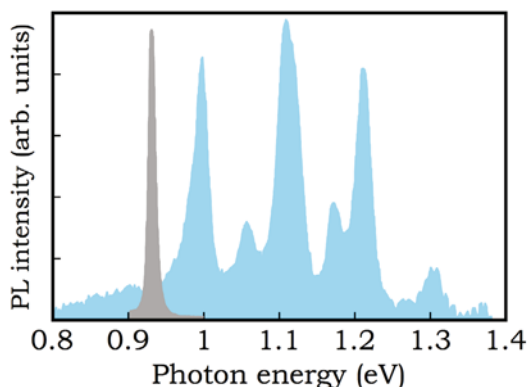


Fig. 2.53 Photoluminescence Spectroscopy. Spectrum of a single walled carbon nanotube (grey) and of an averaged ensemble of carbon nanotubes (cyan) [50]

2.9.7 Auger Spectroscopy

Emission of light is not the only way which the system may de-excite with: the energy released to revert back to the ground state can be employed to kick out one or more electrons. This is named Auger process, and the study of the energy spectrum of the Auger electrons is at the base of Auger Spectroscopy.

2.9.8 Spectroscopies at larger energies

It has been seen that electronic excitations require energies in the range from ultra-violet to X radiation. Other spectroscopies exist that involve higher energy photons and determine either electrons or nuclei excitations. When the photon energy is larger than, say, the electron rest mass or $\simeq 511$ keV, conservation of mass energy and momentum imply that the incoming photon releases only part of its energy to the electron and is then inelastically scattered towards a different direction than the incident one. The peculiar all-quantum situation occurs in which after scattering the original photon decreases its energy and thus increases its wavelength. This is the so called Compton scattering, which occurs more likely between incoming photons with very high energy and almost free electrons, namely electrons of the outermost shells.

Incoming photons with even higher energy, say at least twice the electron rest mass or $\simeq 1022$ keV, may instead interact with the nucleus and give rise to the production of an electron-positron, matter-antimatter, pair, the nucleus absorbing the photon momentum.

2.9.9 Electron Spin Resonance

In Electron Spin Resonance (ESR) measurements, the presence of unpaired electrons can be probed. An unpaired electron in a magnetic field has indeed two possible spin states, corresponding to two different energy values. An incident photon with energy corresponding to the difference in energy between the two states, can be absorbed by the electron and flip its spin.

2.9.10 Neutron scattering

As discussed at the beginning of this section, thermal neutrons are suited probes for magnetic effects in the bulk material. As seen in Chap. 1, elastic scattering of neutrons via their short-range strong interaction with the crystal nuclei can be used to determine the position of nuclei. In Chap. 3 instead is illustrated how inelastic scattering of neutrons via their short-range strong interaction with the crystal nuclei can be used to determine lattice excitations, that are called phonons. On the other hand, neutrons possess a magnetic dipole moment and thus may interact with electronic magnetic moments via dipole-dipole magnetic interactions. Therefore, elastic scattering from thermal neutrons can be used to determine position and orientation of electronic magnetic moments in ferromagnetic and antiferromagnetic samples, whereas inelastic scattering is used to determine the spectrum of magnetic excitations, such as magnons and spin waves. The latter can be measured also via

inelastic scattering with light: this is the Raman scattering method that is discussed in Chap. 3.

Summary: concepts, tools, and procedures to know

Concepts and Tools

- **The electronic structure in crystal and nanostructures can be determined via different methods, each bringing advantages and disadvantages.**
- Tight-Binding (TB) method provides realistic approximations of the electronic structure when electrons are more tightly bound to the corresponding atom in the crystal structure. It worsens otherwise.
- As seen from TB method, the on-site energy E_0 is generally different from the energy of the isolated atom, due to the crystal field. In the evolution from a finite nanocrystal towards an infinite bulk crystal: energy levels thicken inside an energy band with width dictated by first-neighbor overlap integrals, the wavefunction k -vectors thicken inside the range $(0, \pi/a)$, and unoccupied states above the Fermi level become available at no energy cost. Consideration of overlap integrals between states on adjacent sites affects energies and leaves wavefunctions basically unchanged.
- **Particles in the crystal potential can be represented as if they were free particles with an effective mass, or else free quasiparticles.**
- **Real crystals possess surface and impurity states.**
- Whenever the first atom of the chain has an on-site matrix element E'_0 different from the other on-site elements E_0 inside the chain, the formation of a surface state is possible.
- Similar results hold for an impurity state. On general grounds, the conditions for the existence of impurity states are the same as those for surface states.
- The presence of different atomic species in the same chain introduces additional bands, separated by a gap, whose width decreases with increasing the number of cells N .
- **Bloch theorem is a powerful tool to describe electronic structures: allowed and forbidden bands, delocalized and localized states.**
- In a perfect, periodic, infinite crystal, the charge density is periodic with respect to lattice translations. The phase accumulated by the electronic wavefunction after a translation by \mathbf{R}_n is $\mathbf{k} \cdot \mathbf{R}_n$ in terms of lattice vectors \mathbf{R}_n . This is the content of Bloch theorem: $\psi(\mathbf{r} + \mathbf{R}_n) = e^{i\mathbf{k} \cdot \mathbf{R}_n} \psi(\mathbf{r})$.
- Corruption of translational invariance implies complex \mathbf{k} -vectors and therefore localized wavefunctions (such as surface, impurity and other defects states).
- The wavefunction of an electron in a perfect crystal is a plane wave modulated in amplitude by a function periodic on the Bravais lattice. Calculation of this function just in the primitive cell determines the whole electronic wavefunction $\psi(\mathbf{r})$.

- The band energies $E_n(\mathbf{k})$ are even functions of \mathbf{k} , that is $E_n(\mathbf{k}) = E_n(-\mathbf{k})$, and periodic functions on the reciprocal lattice, that is $E_n(\mathbf{k}) = E_n(\mathbf{k} + \mathbf{G})$. This result is verified in all the methods for calculating band structures.
- Gap energies open up at Bragg planes, originated by either differential interactions due to different atomic species in a cell, or as a general quantum-mechanic result of interactions between quantum degenerate states. Within the gap no solutions to the Bloch problem exists: that part of the band is forbidden to electrons.
- The effective masses at a given critical point are in general different for each band and may have different signs.
- Bloch functions can be conveniently constructed starting from sets of atomic wavefunctions, as in Tight Binding method when atomic orbitals are tightly localized close to atoms, and from sets of plane waves, as in Plane Wave method when atomic orbitals are fully delocalized. Midway between the two opposite situations the procedure complexity increases and reliability fails.
- Wannier and Bloch functions are Fourier pairs. Contrary to atomic orbitals, Wannier functions on different sites are orthogonal: in Wannier picture a Bloch function is exactly assembled by taking wavefunctions centered on lattice sites - not to lattice atoms - and connecting them by phases. Energy bands can be constructed from the knowledge of energies at selected locations of the Bravais lattice.
- The Wannier picture works to improve the use of Tight-Binding methods, after upgrading the basis set from incomplete atomic orbitals to complete Wannier functions. Similarly, Pseudopotential methods work to improve the use of Plane-Wave methods, after embodying the fast oscillating part of valence wavefunctions into an effective core: this composes an effective nonlocal potential that acts on effective, now slowly-varying, valence wavefunctions, so that the latter can be expanded with a limited number of plane waves in the basis set.
- **Interactions significantly affect the electronic structure.**
- The effect of electron-nuclei interactions can be handled in most circumstances within the Born-Oppenheimer approximation: due to the slowly-varying positions of heavier nuclei, the electronic eigenvalue problem can be first solved with the nuclei positions \mathbf{R} entering just as parameters.
- The effect of electron-electron interactions can be handled in most circumstances by mapping the system of interacting particles in terms of a system of independent quasi-particles subjected to an effective appropriately constructed single-particle potential, that includes direct local, electrostatic-like, local correlations (Hartree) and exchange nonlocal contribution (Fock).
- In spite of the validity of the original Hartree-Fock approximation, limited to weakly interacting systems where free particle orbitals and Slater nodal structure are reliable, the underlying concept can be extended to account for strong interaction conditions, as detailed in Chap. 6.

- Single-particle energies in Hartree-Fock calculations represent either electron ionization or affinity energies.
- In systems with two electronic states, the states with $N = 1$ (single-electron states) and with $N = 3$ (single-hole states) both behave as if they were single-particle states. States with $N = 2$ behave as excitonic states: they imply energies for promotion of one particle from a lower to a higher energy state, smaller than the energy difference between the two states.
- **Magnetism is a phenomenon due to the interaction between moving charges. Electrons in atoms or molecules composing the material produce small magnetic fields as they spin and move around: thus magnetic effects arise from interactions between spins and moving particles or with external magnetic fields. Magnetic interactions lead to different ground and excited states.**
- Diamagnets are characterized by zero total angular momentum: magnetic moments cancel out in the absence of a magnetic field. When the field is on, the system responds opposing against it via a negative and small magnetic susceptibility. Electrons in outer shells provide the larger contribution to χ_m .
- In paramagnets, χ_m is small as in diamagnets but positive, and scales as the inverse system temperature. Paramagnetism can be very useful for mK cooling, via isentropic cooling processes.
- Exchange interactions between magnetic dipoles may lead to macroscopic magnetic ordering below a critical temperature T_c . This magnetic ordering spontaneously takes place in the absence of an external magnetic field, since it corresponds to a ground state with lower energy than the non-ordered one.
- Ferromagnetic ordering occurs in the case of positive exchange coupling $J > 0$ and can be represented by a lattice of parallel spins, all with the same size of magnetic moment μ . Ferromagnetic ordering occurs below temperatures $T_c = 4N\lambda\mu_b^2S(S+1)/(3k_b)$, often higher than room temperatures. Above T_c the susceptibility behaves as $\chi_m = N\mu^2/[3k_b(T - T_c)]$.
- Antiferromagnetic ordering occurs when $J < 0$ and is characterized by two spin lattices, whose magnetic moments have equal sizes but antiparallel orientation. Antiferromagnetism occurs below the Néel Temperature $T_N = C\mu$ and the susceptibility for $T > T_N$ behaves as $\chi_m = 2C/(T + \theta)$.
- Ferrimagnetic ordering occurs when the spin orientations of two sublattices with different net magnetic moments is parallel. Antiferrimagnetic ordering occurs when the spin orientations of two sublattices with different magnetic moments are antiparallel.
- Magnetically ordered specimen are actually composed of domains, originated from the competition between exchange, anisotropic and magnetic effects, leading to lower energies. Domains are separated by Bloch walls, $\approx a\sqrt{J/K}$ thick and with an energy per unit area $\approx \sqrt{JK/a}$ in terms of the lattice constant a , exchange energy J and anisotropy energy K .

- The lowest-energy normal modes of an infinite linear ferromagnetic spin chain are spin waves, characterized by a quadratic dispersion at long wavelengths. They are called magnons. In antiferromagnets, the $k \rightarrow 0$ dispersion is instead linear.
- **Pairing, attractive, interactions lead to superfluid/ superconducting ground and excited states.**
- Superfluid behavior occurring below a critical temperature, as in, e.g. ^4He with $T_c \simeq 2.2$ K, manifests in linear and rotational flows without dissipation, provided that the linear and angular flow velocities be smaller than critical values. Above these threshold values, sound waves and tiny vortex rings or rotons excitations may set in, where the fluid can be dragged along by the walls and slow down.
- The definition of superfluidity can be related to the resistance that a fluid offers to be twisted by transverse, shear external disturbances and, in second place, to longitudinal compressional perturbations. The superfluid carries no entropy, as if it were one single quantum state, as it occurs in Bose-Einstein Condensation (BEC).
- In BEC, a macroscopic occupation of the lowest energy state occurs below a critical temperature, leading to a sort of condensation of bosonic-like particles but in momentum or energy space. The system behaves as if it were one single particle with well defined momentum and energy. While temperature drops down, the De Broglie wavelength becomes infinitely large, so that the kinetic energy of the lowest state drops to zero and the particle wavefunctions hook together extending over all the volume: they have large probability to be at any point in space.
- BEC and superfluidity are related phenomena. However, the density of Bose-Einstein condensed particles is related to the macroscopic occupation of the lowest energy state, whereas the density of superfluid describes the response of the fluid to a twist, that is a transverse probe. Strong interactions introduce a finite probability that non-condensed particles be carried along with the condensate during the superfluid flow and thus affect the occupation of the lowest energy state depleting the condensate: at ideally zero temperature and in homogeneous systems, the whole system is superfluid but not necessarily all of it is Bose-condensed.
- Superfluidity occurs also in fermionic-like systems, both neutral as in fermionic ^3He isotope or charged as with electrons or holes in superconductors: if a sort of pairing mechanism occurs, the fermion pairs can undergo a form of condensation in momentum space below a T_c , by occupying all one and the same ground state while the Pauli exclusion principle would prevent them to do so. Key features of superconductors are zero-resistance behavior, transparency to radiation with frequency below a critical ω_c , exponential behavior of specific heat and thus for activation of thermal excitations, and perfect diamagnetism below T_c .

- Superconductivity or superfluidity of both bosonic and fermionic-like systems are related phenomena, they all being based on the existence of Bose-Einstein condensed state with wavefunction $\psi(\mathbf{r}) = \sqrt{n_c(\mathbf{r})}e^{i\phi(\mathbf{r})}$ in terms of a condensate density $n_c(\mathbf{r})$ and a global space-dependent phase $\phi(\mathbf{r})$. The superfluid velocity v_s is dictated by the space gradient of the phase. In a fermion system, the condensate wavefunction becomes the pairing wavefunction $\psi(\mathbf{r}) \rightarrow p(\mathbf{r}) \equiv \langle \psi_{\downarrow}(\mathbf{r})\psi_{\uparrow}(\mathbf{r}) \rangle$, and corresponding to the situation in which the Bose particle (integer spin) is represented by two paired fermions with opposite spins. Other combinations leading to integer spin are possible, and fermion-pairs wavefunctions do spatially overlap.
- The BCS theory explains superconductivity by assuming that electrons close to the Fermi level are correlated in pairs, so that an energy gap $\Delta(T)$ exists between the ground and the first excited single particle states essentially related to the binding energy of the pair. Since the pairs are highly overlapped in real space though, the gap energy does not coincide with the pair binding energy.
- The dependence of T_c on isotopic species and on vibrational energy supports the idea that the pairing mechanism in conventional superconductors is originated by lattice vibrations. Other pairing mechanisms are expected to be involved in different superconductors, such as the high- T_c ones.
- Whatever the microscopic mechanism might be for the pairing, the BCS theory has shown to be a general framework to treat superconductivity and superfluidity phenomena involving fermionic-like particles, and can be demonstrated to continuously evolve towards the theory of Bose-Einstein Condensation for point-like bosonic particles.
- **Probes to determine the electronic structure include either photons or massive particles with energies above the 0.1 to 10 eV range of the electronic excitations under study, and wavelengths comparable to Å interparticle distances, as well as the investigation of transport properties faced in Chap. 4.**
- The massive particles may in turn be either charged or neutral, such as electrons or neutrons. on general grounds, charged particles experience strong, long-ranged interactions while entering the material: therefore they are able to penetrate only for a few layers and work as surface-sensitive probes. Neutrons are uncharged and the strong-force interaction which they act with is very short-ranged: scattering processes based on neutrons thus work as bulk probes.
- Electron scattering is suited to probe high-energy excitations such as collective electronic excitations and interband transitions, whereas thermal neutrons are most suited to investigate low-energy excitations such as the vibrational modes of the crystal or the spin-wave, magnon modes in ferro and antiferromagnets.

Procedures

- Use of the Semi-empirical Tight-Binding method to determine energies and wavefunctions in 1D nanostructures with given boundary conditions

- Use of Bloch theorem and approximated periodic boundary conditions to determine the electronic structure of a finite crystal, reducing the original problem complexity, which results in a $(NN_b) \times (NN_b)$ sized matrix.
 - Fold band structures into the first Brillouin zone.
 - Minimize energy functionals, as in Hartree and Hartree-Fock methods.
 - Calculate electron ionization and affinity energies.
 - Calculate nonlocal exchange terms (Slater procedure).
 - Determination of electronic spectrum of nanostructures.
-

2.10 Appendix. Elements of variational calculus

A real function of real variables in one dimension is an univocal correspondence: to each given value of one independent variable, one and only one real number is associated according to a given rule. In more than one dimensions (2D or 3D), the independent variable is replaced by a set of independent variables, e.g. two or three. A functional is instead a correspondence where the variable that is made independently varying is a function, instead than a number. Thus, a functional is a correspondence between a function and a real number, generated according to a given rule. Considering the type of the variational calculations one has to execute, the needed useful tools can be introduced by means of the following simple examples.

First, a few examples of functionals. The object $F_1[f]$ defined as

$$F_1[f] = \int_a^b dx h(x) f^2(x) \quad (2.273)$$

is a functional, since it produces a number for each function $f(x)$, once the function $h(x)$ is known. Note that the argument of the functional is given between square brackets $[]$, so that round brackets $()$ can be reserved to arguments of ordinary real functions such as in $f(x)$. The object

$$F_2[f] = \int_a^b dx g(x) \left[\frac{df(x)}{dx} \right]^2, \quad (2.274)$$

is also a functional, since it produces a number for each derivative of $f(x)$, again once $g(x)$ is given. And so on.

Second, the variation δF of a functional $F[f]$ with respect to a function $f(x)$ has to be defined. This is naturally defined in analogy with variations of functions with respect to their argument, that is:

$$\delta F = F[f + \delta f] - F[f], \quad (2.275)$$

where $\delta f(x)$ is an arbitrary function summed to $f(x)$. Apply this definition to (2.273). One has:

$$\begin{aligned}
\delta F_1 &= \int_a^b dx h(x) [f(x) + \delta f(x)]^2 - \int_a^b dx h(x) f^2(x) \\
&= \int_a^b dx h(x) [2f(x)\delta f(x) + [\delta f(x)]^2] .
\end{aligned} \tag{2.276}$$

Notice that one term appears in $\delta f(x)$ and one in $(\delta f(x))^2$. If one sets $\delta f(x) \rightarrow \varepsilon \delta f(x)$ and performs the $\varepsilon \rightarrow 0$ limit, one would find first and second order terms. Neglecting higher-order terms, one has:

$$\delta F_1 = 2 \int_a^b dx h(x) f(x) \delta f(x) , \tag{2.277}$$

that is the expression typically used in the text for the variation of functionals like δF_1 . Following the same steps, for functionals like $F_2[f(x)]$ instead one obtains:

$$\begin{aligned}
\delta F_2 &= \int_a^b dx g(x) \left[\frac{d}{dx} (f(x) + \delta f(x)) \right]^2 - \int_a^b dx g(x) \left(\frac{df(x)}{dx} \right)^2 \\
&= \int_a^b dx g(x) \left[2 \frac{df(x)}{dx} \frac{d\delta f(x)}{dx} + \left(\frac{d\delta f(x)}{dx} \right)^2 \right] \\
&= \int_a^b dx g(x) 2 \frac{df(x)}{dx} \frac{d\delta f(x)}{dx} ,
\end{aligned} \tag{2.278}$$

where terms higher than first order have been again neglected. Expression (2.278) is however not easily handled yet, since it contains the derivative of function $\delta f(x)$. An integration by parts can resolve this problem, yielding

$$\delta F_2 = \left(2g(x) \frac{df(x)}{dx} \delta f(x) \right)_a^b - \int_a^b dx \frac{d}{dx} \left(2g(x) \frac{df(x)}{dx} \right) \delta f(x) .$$

that is amenable to much easier handling. When the integration limits are given as $\delta f(a) = \delta f(b) = 0$, one finds

$$\delta F_2 = - \int_a^b dx \frac{d}{dx} \left(2g(x) \frac{df(x)}{dx} \right) \delta f(x) , \tag{2.279}$$

that is the expression typically used in the text for the variation of functionals like δF_2 . The above rules to calculate δF are easily extended to cases in which the exponent in (2.273) and (2.274) is a generic integer n other than 1 or 2.

In the examples discussed it results that

$$\delta F[f] = \int_a^b dx G(x) \delta f(x) ,$$

but this occurs for any functional. It is possible therefore to define the functional derivative in the following way: given $\delta F[f]$ and $\delta f(x) = \delta f(x_0) \delta(x - x_0)$, the derivative of the functional in $f(x_0)$ is

$$\frac{\delta F[f]}{\delta f(x_0)} = \int_a^b dx G(x) \delta(x - x_0) = G(x_0)$$

2.10.1 Variational theorem

A third useful tool entering variational calculus is related to the concept of stationary conditions for the functional $F[f]$: $F[f]$ is stationary if $\delta F = 0$ whatever the variation of the argument $\delta f(x)$ might be. Reverting back to an example, suppose to be in the mood of determining stationary conditions of the functional

$$F[f] = \int_a^b dx \left\{ h(x) f^2(x) + g(x) \left[\frac{df(x)}{dx} \right]^2 \right\}, \quad (2.280)$$

given the conditions $\delta f(a) = \delta f(b) = 0$ and with the following additional constraint on $f(x)$:

$$\int_a^b dx f^2(x) - 1 = 0. \quad (2.281)$$

It can be shown that Lagrange theorem to calculate stationary points of multiple-variables functions with additional constraints still holds. Stationary conditions are thus determined after considering the functional variation

$$\bar{F}[f] = \int_a^b dx \left[h(x) f^2(x) + g(x) \left(\frac{df(x)}{dx} \right)^2 \right] - \lambda \left(\int_a^b dx f^2(x) - 1 \right), \quad (2.282)$$

with λ a constant parameter, referred to as Lagrange multiplier. One then finds

$$\delta \bar{F} = \int_a^b dx \delta f(x) \left[2h(x)f(x) - 2\frac{d}{dx} \left(g(x) \frac{df(x)}{dx} \right) - 2\lambda f(x) \right]. \quad (2.283)$$

Thus necessary and sufficient condition to get $\delta \bar{F} = 0$ for whatever variation $\delta f(x)$ is:

$$h(x)f(x) - \frac{d}{dx} \left(g(x) \frac{df(x)}{dx} \right) - \lambda f(x) = 0. \quad (2.284)$$

If $g(x)$ were a constant, as for example is $\hbar^2/(2m)$, (2.284) becomes

$$-\frac{\hbar^2}{2m} \frac{d^2 f(x)}{dx^2} + h(x)f(x) = \lambda f(x), \quad (2.285)$$

that is the Schrödinger equation for a particle in potential $h(x)$.

This is indeed a crucial result: the stationary condition on the functional

$$F[f] = \int_a^b dx \left[h(x) f^2(x) + \frac{\hbar^2}{2m} \left(\frac{df(x)}{dx} \right)^2 \right], \quad (2.286)$$

with the additional constraint (2.281) is equivalent to the Schrödinger equation (2.285), in the sense that the statement holds in both directions.

In the case of the lowest-energy or fundamental state, the following procedure applies. Try a given function $f_0(x, \alpha, \beta, \dots)$, from now on referred to as trial wave function, normalized to unity and depending on a number of parameters α, β, \dots . The functional (2.286) becomes now a function of those parameters $F(\alpha, \beta, \dots)$, and gets its minimum value when $\partial F / \partial \alpha = \partial F / \partial \beta = \dots = 0$. Be α_m, β_m, \dots the values of α, β, \dots which satisfy the minimum condition. Thus one has

$$\int_a^b dx \left[h(x) f_0^2(x, \alpha_m, \beta_m, \dots) + \frac{\hbar^2}{2m} \left(\frac{df_0(x, \alpha_m, \beta_m, \dots)}{dx} \right)^2 \right] \geq \lambda_0, \quad (2.287)$$

with λ_0 being the lowest eigenvalue of (2.285).

The above theorem has actually a very general character. It includes for example the case in which variable x represents a spatial coordinate \mathbf{r} . In this case the starting functional is

$$F[f] = \int_V d\mathbf{r} \left[h(\mathbf{r}) f^2(\mathbf{r}) + \frac{\hbar^2}{2m} (\nabla f(\mathbf{r}))^2 \right], \quad (2.288)$$

with V the space volume where the problem is considered. If condition

$$\int_V d\mathbf{r} f^2(\mathbf{r}) - 1 = 0 \quad (2.289)$$

has to be satisfied yet and $f(\mathbf{r})$ vanishes at the given volume boundaries, the stationary condition assumes the form

$$-\frac{\hbar^2}{2m} \nabla^2 f(\mathbf{r}) + h(\mathbf{r}) f(\mathbf{r}) = \lambda f(\mathbf{r}). \quad (2.290)$$

The stationary-condition theorem holds as well if V represents the whole space or if the function $f(\mathbf{r})$ is complex-valued. In the latter case the functional to be considered is

$$F[f] = \int d\mathbf{r} \left[h(\mathbf{r}) |f(\mathbf{r})|^2 + \frac{\hbar^2}{2m} |\nabla f(\mathbf{r})|^2 \right], \quad (2.291)$$

and the condition (2.289) becomes

$$\int d\mathbf{r} |f(\mathbf{r})|^2 - 1 = 0. \quad (2.292)$$

2.11 Appendix. Calculation of matrix elements within the Hartree-Fock approximation

Consider (2.137). It is seen that H_e can be expressed as

$$\begin{aligned}
H_e &= A_0 + A_1 + A_2, \\
A_0 &= \frac{1}{2} \sum_{I,J,(I \neq J)} \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|} \\
A_1 &= \sum_i \left[-\frac{\hbar^2}{2m} \nabla_i^2 - \sum_I \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} \right] \\
A_2 &= \frac{1}{2} \sum_{i,j(i \neq j)} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{2.293}
\end{aligned}$$

Here, A_0 is an operator not containing the electronic variables, A_1 the sum of electronic single-particle operators, and A_2 the sum of electronic operators containing variables related to pairs of particles.

It is also known that determinants can be expanded as a sum of $N!$ terms, each of them being composed of the product of different elements with i and j indexes as follows: the first index can be expressed as a fundamental permutation and the second index is one out of $N!$ possible indexes permutations. The resulting term sign is positive for even numbers of exchanges, and negative for odd ones.

In calculating the matrix elements of the operators A_0 and A_1 , only the terms possessing the same permutation of second indexes are non zero: this occurs because the integral of rank $3N$ becomes the product of N integrals of rank 3, each of them yielding the value 1. If the permutation of the indexes is not the same, at least one integral is zero. For this reason, the matrix element $\langle \Phi | A_0 | \Phi \rangle = A_0$ and that of $\langle \Phi | A_1 | \Phi \rangle$ is

$$\langle \Phi | A_1 | \Phi \rangle = \sum_{i=1}^N \int d\mathbf{r} \psi_i^*(\mathbf{r}) \left[-\frac{\hbar^2}{2m} \nabla^2 - \sum_{I=1}^M \frac{Z_I e^2}{|\mathbf{r} - \mathbf{R}_I|} \right] \psi_i(\mathbf{r}). \tag{2.294}$$

Non-vanishing matrix elements of a single-particle operator are indeed those possessing the same permutation of second indexes in the two determinants.

In the calculation of non-vanishing matrix elements of the type $\langle \Phi | A_2 | \Phi \rangle$ instead, only $N - 2$ permutations are to be identical in the two determinants, so that the two other indexes are either the same or swapped. One thus obtains:

$$\begin{aligned}
\langle \Phi | A_2 | \Phi \rangle &= \frac{e^2}{2} \sum_{i,j=1}^N \left[\int d\mathbf{r} d\mathbf{r}' \frac{|\psi_i(\mathbf{r})|^2 |\psi_j(\mathbf{r}')|^2}{|\mathbf{r} - \mathbf{r}'|} \right. \\
&\quad \left. - \delta_{m_i, m_j} \int d\mathbf{r} d\mathbf{r}' \frac{\psi_i^*(\mathbf{r}) \psi_i(\mathbf{r}') \psi_j(\mathbf{r}) \psi_j^*(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \right], \tag{2.295}
\end{aligned}$$

where the term δ_{m_i, m_j} gets the value 1 if states φ_i and φ_j share the same spin function, and the value 0 otherwise.

2.12 Appendix. Electronic bands calculation in the jellium model

Consider the exchange term in (2.147):

$$\begin{aligned}
 U_{xc} &= -e^2 \sum_{j=1}^N \delta_{m_i, m_j} \int \frac{\psi_i(\mathbf{r}') \psi_j(\mathbf{r}) \psi_j^*(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \\
 &= -e^2 \sum_{j=1}^N \delta_{m_i, m_j} \frac{1}{V^{\frac{3}{2}}} \int d\mathbf{r}' \frac{e^{i\mathbf{k}_j \cdot (\mathbf{r} - \mathbf{r}')} e^{i\mathbf{k}_i \cdot \mathbf{r}'}}{|\mathbf{r} - \mathbf{r}'|} \\
 &= -e^2 \sum_{j=1}^N \delta_{m_i, m_j} \frac{1}{V^{\frac{3}{2}}} \int d\mathbf{r}' \frac{e^{i\mathbf{k}_j \cdot (\mathbf{r} - \mathbf{r}')} e^{-i\mathbf{k}_i \cdot (\mathbf{r} - \mathbf{r}')}}{|\mathbf{r} - \mathbf{r}'|} V^{\frac{1}{2}} \frac{1}{V^{\frac{1}{2}}} e^{i\mathbf{k}_i \cdot \mathbf{r}} \\
 &= -e^2 \sum_{j=1}^N \delta_{m_i, m_j} \frac{1}{V} \int d\mathbf{r}' \frac{e^{i(\mathbf{k}_j - \mathbf{k}_i) \cdot (\mathbf{r} - \mathbf{r}')}}{|\mathbf{r} - \mathbf{r}'|} \psi_i(\mathbf{r}) . \tag{2.296}
 \end{aligned}$$

In this case the exchange potential results to be local, and the corresponding value is

$$V_{xc} = -e^2 \sum_{j=1}^N \delta_{m_i, m_j} \frac{1}{V} \int d\mathbf{r}' \frac{e^{i(\mathbf{k}_j - \mathbf{k}_i) \cdot (\mathbf{r} - \mathbf{r}')}}{|\mathbf{r} - \mathbf{r}'|} . \tag{2.297}$$

The integral in (2.297) is the Fourier transform of the $1/r$ -like Coulomb potential at point $\mathbf{k}_j - \mathbf{k}_i$. Besides, the sum over j can be conveniently substituted by an integral according to the correspondence $\sum_j \rightarrow V(2\pi)^{-3} \int d\mathbf{k}_j$. With these tricks one eventually obtains

$$V_{xc}(k) = -\frac{4\pi e^2}{(2\pi)^3} \int_{q < k_f} d\mathbf{q} \frac{1}{|\mathbf{k} - \mathbf{q}|^2} , \tag{2.298}$$

where the replacements $\mathbf{k}_i \rightarrow \mathbf{k}$ and $\mathbf{k}_j \rightarrow \mathbf{q}$ have been set.

The integral can be calculated in polar coordinates, choosing the z -axis direction along \mathbf{k} . One has:

$$\begin{aligned}
 V_{xc}(k) &= -\frac{4\pi e^2}{(2\pi)^3} \int_{q < k_f} \frac{1}{q^2 + k^2 - 2kq \cos \theta} q^2 \sin \theta d\theta d\varphi dq \\
 &= -\frac{e^2}{\pi} \int_{q < k_f} \frac{q^2}{q^2 + k^2 - 2kq \cos \theta} d(-\cos \theta) dq \\
 &= -\frac{e^2}{\pi k} \int_0^{k_f} q \ln \left| \frac{k+q}{k-q} \right| dq . \tag{2.299}
 \end{aligned}$$

This last integral can be analytically calculated, yielding the final result

$$V_{xc}(k) = -\frac{2e^2 k_F}{\pi} g\left(\frac{k}{k_F}\right) ,$$

$$g(x) = \frac{1}{2} + \frac{1-x^2}{4x} \ln \frac{|1+x|}{|1-x|}. \quad (2.300)$$

2.13 Appendix. Bose-Einstein condensation: basic concepts and more recent realizations

Quantum behavior of bosonic particles is ruled by Bose-Einstein statistics. This states that the occupation number of a state with energy ε is given by

$$n(\varepsilon) = \frac{1}{e^{\beta(\varepsilon-\mu)} - 1}, \quad (2.301)$$

with $\beta = (k_b T)^{-1}$. Quasi-free and weakly interacting particles are characterized by a continuous band of positive energies. Therefore the chemical potential is to be negative, to prevent that the distribution function (2.301) have a singularity. This is a crucial difference with respect to weakly interacting fermions ruled by the Fermi-Dirac statistics, that are characterized by positive μ values. This difference has relevant consequences. Assume for simplicity to have a bosonic system of free particles with $g(\varepsilon)$ the density of states per unit volume. The relation connecting the number of particles to the chemical potential is

$$N = V \int_0^\infty d\varepsilon \frac{g(\varepsilon)}{e^{\beta(\varepsilon-\mu)} - 1}. \quad (2.302)$$

While temperature T drops down, the de Broglie wavelength $\lambda_{dB} = \sqrt{2\pi\hbar^2/(mk_b T)}$ of the particles with mass m progressively extends to get a macroscopic size, the kinetic energy of the system progressively dropping to vanishing values. In the meanwhile, conservation of the number of particles requires that μ increases from larger negative values to vanishing values. At the temperature, say T_c , at which $\mu = 0$, the quantity

$$N' = V \int_0^\infty d\varepsilon \frac{g(\varepsilon)}{e^{\beta\varepsilon} - 1} \quad (2.303)$$

has the property that $N' < N$, and conservation of the number of particles can be accomplished only if a macroscopic occupation N_0 of the ground state occurs so that

$$N(T) = N_0(T) + N'(T).$$

The N_0 particles are condensed in momentum space, that is they behave coherently as if they were one single particle with the same momentum. In the ground state with vanishing kinetic energy and in the absence of external forces, this momentum is zero. At zero temperature, all the N particles are expected to be momentum-condensed in the ground state so that $N(T=0) = N_0$. In this sense they represent a different state of matter than solid, liquid, or gaseous forms. The manifestation of such a state of matter process has been predicted at the beginning of the 20th

century by Bose and Einstein to occur in photon and atomic systems, respectively, and is named Bose-Einstein condensation. T_c is named critical temperature.

When interactions are switched on, particles can be excited out of the condensate even at zero temperature. According to an elegant argument given by Feynman, since the bosonic wavefunction is symmetrical for exchange of particles even at very long distances, such an excitation requires a vanishingly small energy and is therefore a long-wavelength density fluctuation: to such an excitation the name of acoustic phonon is given in Chap. 3. Interactions may cause non-condensed particles to be carried along with the condensate and thus affect the occupation of the lowest energy state: they deplete the condensate, while the latter keeps a macroscopic occupation $N_0/N \ll 1$ though. In this sense, a Bose liquid is well behaved [51] in much the same way as a normal Fermi liquid is: irrespective of interactions, there always is a well-defined Fermi surface where to drive excitations from. In Bose liquids, the condensate plays the role that the Fermi sphere has in Fermi liquids. The macroscopic coherent behavior remains, that leads to the observation of the many peculiar manifestations of liquid helium physics, first of all superfluid motion, that is motion with no energy dissipation.

For many years the quantum liquid of bosonic-like helium ^4He particles has been the best known system where Bose-Einstein condensation occurs at critical temperatures of a few Kelvin above absolute zero: in liquid helium though, particles in the quantum fluid are strongly interacting and therefore a significant quantum depletion of the ground state occurs along the mechanisms illustrated above, at the level of 90%. Bose-Einstein condensation of excitons in solids has been investigated as an example of Bose condensation for weakly interacting particles, more recently reaching the experimental demonstration. The first striking observation of Bose-Einstein condensation of almost non-interacting particles has been realized in 1995 in the groups of Wiemann and Cornell at JILA [52], and Ketterle at MIT [53]: a few thousands of bosonic isotopes of alkali atoms were cooled down to a few tens of nK by a combined technique of magneto-optical trapping, laser and evaporative cooling. Magneto-optical trapping serves to increase the system density without suffering dissipative processes against physical walls. Laser cooling is then used for a first slowing down of atoms along the three space directions. Eventually, the most effective evaporative cooling is applied: hotter atoms are expelled from the trap by radiofrequency transitions and the remaining ones thermalize to lower temperatures by elastic collisions. During the different cooling processes down to temperatures extraordinarily close to absolute zero, atoms are kept in a metastable gaseous dilute state with density n . The description of atomic interactions in these systems is amenable to a very accurate modeling by means of a contact, δ -like pseudopotential $U(r) = 4\pi\hbar^2 a \delta(\mathbf{r})/M$ driven by one single parameter that is the scattering length a , besides the atom mass M . The scattering length can be measured in independent experiments. The diluteness condition $na^3 \ll 1$ implies a weak interacting regime, at least in the original experiments.

Atomic physics systems as those where Bose-Einstein Condensation has been realized, are actually characterized by the possibility of very accurate manipulation of atoms and their internal and external interactions, and of accurate related mea-

surements [54]. In fact, a number of different confinement geometries have been realized, from 3D down to 0D confinements. To this aim, periodic lattice-like potentials are used composed of counterpropagating beams of stationary laser light detuned from the atomic transitions, where the atoms can surf on or tunnel through, the lattice parameter being half the laser wavelength $\lambda_L/2$: one optical lattice realizes regular sets of 2D sheets of atomic samples, two orthogonal lattices realize regular sets of 1D tubes, three orthogonal lattices realized regular sets of 0D egg-like dots. In many realizations the scattering length a can be tuned by means of external fields to be any value from very large and negative to very large and positive, passing through zero, thus providing a tool to drive the same system from attractive to repulsive or no interactions at all. Use of atomic species with large dipolar momentum offers a way to tune the range of the interactions as well. Last and not least, atoms with fermionic statistics can be trapped as well and in fact Bose-Einstein condensation and superfluidity of pairs of fermions has been predicted and observed.

In essence, quantum gases represent quite interesting systems for fundamental physics and applications, where extreme conditions of quantum degeneracy and correlations can be reached and tuned almost at will, by changing temperature and tuning interactions, using reduced dimensions and different quantum statistics. Applications are on the way, such as in the fields of quantum information and computing, quantum transport and atomtronic realizations, and fundamental physics as in tests of generality relativity and quantum gravitation.

2.14 Appendix. Representation of many particle fermion and boson states by means of the occupation numbers

The present Appendix reminds the basic tools to manage quantum operators for fermion and boson states, useful in the rest of this textbook.

2.14.1 Fermion states

The state of N interacting fermions can be described after starting from a complete base of single-particle states $\varphi_n(q)$ combined into the Slater determinant

$$\Phi(q_1, q_2, \dots, q_N) = \text{Det} \begin{pmatrix} \varphi_1(q_1) & \varphi_1(q_2) & \dots & \varphi_1(q_N) \\ \varphi_2(q_1) & \varphi_2(q_2) & \dots & \varphi_2(q_N) \\ \dots & \dots & \dots & \dots \\ \varphi_N(q_1) & \varphi_N(q_2) & \dots & \varphi_N(q_N) \end{pmatrix}. \quad (2.304)$$

The many-fermions state is represented by a linear combination of such determinants. The state can be determined after indicating the number of particles in each state. This is to be accomplished by keeping into accounts the Pauli exclusion prin-

ciple, stating that no more than one particle can be in a given state. In fact, if it were not so, two particles could have the same wavefunction and the determinant (2.304) would be zero.

The state representing the fermion system can be cast in the form $\Psi(n_1, n_2, \dots, n_i, \dots)$, where the numbers n_1, n_2, \dots, n_i label as many particles - 1 or 0 for fermions - are in the states $1, 2, \dots, i, \dots$. The creation and destruction operators are now defined as

$$\begin{aligned} c_i^\dagger \Psi(n_1, n_2, \dots, n_i, \dots) &= \theta_i (1 - n_i) \Psi(n_1, n_2, \dots, 1 - n_i, \dots) \\ c_i \Psi(n_1, n_2, \dots, n_i, \dots) &= \theta_i n_i \Psi(n_1, n_2, \dots, 1 - n_i, \dots), \end{aligned} \quad (2.305)$$

with $\theta_i = (-1)^{v_i}$ and $v_i = \sum_{p=1}^{i-1} n_p$. The quantity $\theta_i = \pm 1$, according to whether the sum of the occupation numbers of the states preceding the state i is even (+ sign) or odd (- sign). The fermion operators satisfy the anticommutation relations

$$\begin{aligned} \{c_i, c_j^\dagger\} &= c_i c_j^\dagger + c_j^\dagger c_i = \delta_{i,j} \\ \{c_i, c_j\} &= c_i c_j + c_j c_i = 0 \\ \{c_i^\dagger, c_j^\dagger\} &= c_i^\dagger c_j^\dagger + c_j^\dagger c_i^\dagger = 0. \end{aligned} \quad (2.306)$$

It can be easily shown that

$$c_i^\dagger c_i \Psi(n_1, n_2, \dots, n_i, \dots) = n_i \Psi(n_1, n_2, \dots, n_i, \dots). \quad (2.307)$$

In essence, the operator $c_i^\dagger c_i$ destroys and then creates one fermion in state i , that is to say that the operator $c_i^\dagger c_i$ counts the number of particles in the state i .

2.14.2 Boson states

Consider first the harmonic oscillator described by the Hamiltonian

$$H = \frac{1}{2M} p^2 + \frac{M}{2} \omega^2 x^2. \quad (2.308)$$

The corresponding eigenvalues and eigenvectors are:

$$\begin{aligned} E_n &= \hbar \omega \left(n + \frac{1}{2} \right), \\ \psi_n(x) &= H_n(M\omega x) e^{-\frac{M^2 \omega^2}{2} x^2}, \end{aligned} \quad (2.309)$$

with n a positive integer and $H_n(M\omega x)$ the Hermite polynomial of order n . The operators

$$\begin{aligned}
b^\dagger &= \sqrt{\frac{M\omega}{2\hbar}}x - i\sqrt{\frac{1}{2\hbar M\omega}}p, \\
b &= \sqrt{\frac{M\omega}{2\hbar}}x + i\sqrt{\frac{1}{2\hbar M\omega}}p
\end{aligned} \tag{2.310}$$

satisfy the Bose commutation relations

$$[b, b^\dagger] = 1. \tag{2.311}$$

The Hamiltonian H can be expressed in terms of b and b^\dagger as

$$H = \hbar\omega \left(b^\dagger b + \frac{1}{2} \right). \tag{2.312}$$

Using the notation $|n\rangle$ for the state $\psi_n(x)$ and $|0\rangle$ for the ground state, one has that $b|0\rangle = 0$, that is

$$\left(\sqrt{\frac{M\omega}{2\hbar}}x + i\sqrt{\frac{1}{2\hbar M\omega}}p \right) \psi_0(x) = 0. \tag{2.313}$$

Using now the recurrence relations between Hermite polynomials, one also obtains:

$$\begin{aligned}
b^\dagger|n\rangle &= \sqrt{n+1}|n+1\rangle, \\
b|n\rangle &= \sqrt{n}|n-1\rangle,
\end{aligned} \tag{2.314}$$

that is

$$\begin{aligned}
\left(\sqrt{\frac{M\omega}{2\hbar}}x - i\sqrt{\frac{1}{2\hbar M\omega}}p \right) \psi_n(x) &= \sqrt{n+1}\psi_{n+1}(x), \\
\left(\sqrt{\frac{M\omega}{2\hbar}}x + i\sqrt{\frac{1}{2\hbar M\omega}}p \right) \psi_n(x) &= \sqrt{n}\psi_{n-1}(x), \\
b^\dagger b|n\rangle &= n|n\rangle.
\end{aligned} \tag{2.315}$$

The excited states of the harmonic oscillator can be found by successive application of the operator b^\dagger onto the state $|0\rangle$. One indeed finds that

$$|n\rangle = \frac{1}{\sqrt{n!}} (b^\dagger)^n |0\rangle. \tag{2.316}$$

In essence, the state $|n\rangle$ can be obtained by successive creation of n quanta with energy $\hbar\omega$.

As in the case of fermions, the state of N bosons can be described after starting from a complete base of single-particle states $\phi_n(q)$ and considering a linear combination of functions of the type

$$\Phi(q_1, q_2, \dots, q_N) = P \begin{pmatrix} \varphi_1(q_1) & \varphi_1(q_2) & \dots & \varphi_1(q_N) \\ \varphi_2(q_1) & \varphi_2(q_2) & \dots & \varphi_2(q_N) \\ \dots & \dots & \dots & \dots \\ \varphi_N(q_1) & \varphi_N(q_2) & \dots & \varphi_N(q_N) \end{pmatrix} \quad (2.317)$$

where the indexes $1, 2, \dots, N$ label N states of the set and q_1, q_2, \dots, q_N the coordinates of each particle, spin included. The symbol P to the right hand of the previous equation means that, at variance with the fermion case, the quantity is obtained by calculating the determinant with the usual rule except that this time the signs are to be all positive. Yet at variance with the fermion case, the wavefunction (2.317) is symmetric when two particles are exchanged and it can also be represented indicating the number of particles that are in a given single-particle state. Therefore, the state representing the boson system can be cast in the form $\Psi(n_1, n_2, \dots, n_i, \dots)$, where the numbers n_1, n_2, \dots, n_i label as many particles as the states $1, 2, \dots, i, \dots$. The creation and destruction operators are defined as

$$\begin{aligned} b_i^\dagger \Psi(n_1, n_2, \dots, n_i, \dots) &= \sqrt{n_i + 1} \Psi(n_1, n_2, \dots, n_i + 1, \dots) \\ b_i \Psi(n_1, n_2, \dots, n_i, \dots) &= \sqrt{n_i} \Psi(n_1, n_2, \dots, n_i - 1, \dots). \end{aligned} \quad (2.318)$$

They satisfy the commutation relations

$$\begin{aligned} [b_i, b_j^\dagger] &= b_i b_j^\dagger - b_j^\dagger b_i = \delta_{i,j} \\ [b_i, b_j] &= b_i b_j - b_j b_i = 0 \\ [b_i^\dagger, b_j^\dagger] &= b_i^\dagger b_j^\dagger - b_j^\dagger b_i^\dagger = 0. \end{aligned} \quad (2.319)$$

Furthermore, one has that

$$b_i^\dagger b_i \Psi(n_1, n_2, \dots, n_i, \dots) = n_i \Psi(n_1, n_2, \dots, n_i, \dots). \quad (2.320)$$

That is to say that, as in the fermion case, $b_i^\dagger b_i$ indeed counts the number of bosons in state i .

Problems with solutions

2.1. Given the linear equations

$$\begin{aligned} E_0 \phi_n + \gamma(\phi_{n+1} + \phi_{n-1}) &= E \phi_n \text{ with } n = 2, \dots, N, \\ (E'_0 - E) \phi_1 + \gamma \phi_2 &= 0, \end{aligned} \quad (2.321)$$

and the condition $\phi_{N+1} = 0$, show that $\phi_n = A \sin(nka + \delta)$ satisfies them if

$$E = E_0 + 2\gamma \cos ka,$$

$$\tan \delta = -\frac{(E'_0 - E_0) \sin ka}{(E'_0 - E_0) \cos ka - \gamma}$$

$$ka = \frac{\nu\pi - \delta}{N+1} . \quad (2.322)$$

Solution. Substitute ϕ_n in the first and the second line of the (2.321) and obtain the first and the second line of (2.322). The condition $ka = (\nu\pi - \delta)(N+1)$ is originated by the boundary condition $\phi_{N+1} = 0$.

2.2. A 1D lattice is modeled by potentials of the form $V_0 a \delta(x)$, with a the lattice parameter. Determine the Fourier coefficients of the periodic potential $V(x) = V_0 a \sum_n \delta(x - na)$. How large are the gaps with this model, if the electrons are considered independent?

Solution. By definition $U(G) = \frac{1}{a} \int_{-a/2}^{a/2} V(x) dx e^{-iGx}$, with $G = 2\pi p/a$ and p integer. One has then $U(G) = a^{-1} \int_{-a/2}^{a/2} dx V_0 a \delta(x) e^{-iGx} = V_0$. All the gaps between the different bands result to have the same value.

2.3. Consider the model $V_0 a \delta(x)$ for the potential of a single atom. Calculate the band structure of a 1D crystal with lattice parameter a by means of the Tight-Binding method. Perform the calculation twice: for the case of an infinite crystal including overlaps up to second-neighbors and for a finite crystal including only first-neighbor overlaps.

Solution. The wavefunction of one single atom is solution of the Schrödinger equation

$$-\frac{\hbar^2}{2m} \varphi''(x) + V_0 a \delta(x) \varphi(x) = \varepsilon_0 \varphi(x) .$$

In the two regions $x > 0$ and $x < 0$, the solution $\varphi(x)$ can be expressed as:

$$\varphi(x) = \begin{cases} A e^{-\kappa x} & \text{for } x > 0 \\ A e^{\kappa x} & \text{for } x < 0 \end{cases} ,$$

with $\kappa = \sqrt{2m\varepsilon_0}/\hbar$ in order to have an integrable squared function. Wavefunction continuity has to be guaranteed, so that the constant A is the same for $x \rightarrow 0^\pm$. One has:

$$\varphi'(0^+) - \varphi'(0^-) = \frac{2m}{\hbar^2} V_0 a \varphi(0) ,$$

that is $\kappa = -mV_0 a/\hbar^2$. The eigenvalue ε_0 can thus be found. To have a positive value of κ , it must be $V_0 < 0$. The normalization condition leads to $A = \sqrt{\kappa}$. In order to solve now the eigenvalue problem

$$-\frac{\hbar^2}{2m} \psi''(x) + V_0 a \sum_n \delta(x - na) \psi(x) = \varepsilon \psi(x) , \quad (2.323)$$

one can write

$$\psi(x) = \sum_p \phi_p \varphi(x - pa) . \quad (2.324)$$

By substituting (2.324) into (2.323), multiplying by $\varphi(x - qa)$ and integrating over, one gets

$$\sum_p \phi_p \{ (\varepsilon_0 - \varepsilon) s_{q,p} + U_{q,p} \} = 0 , \quad (2.325)$$

with

$$s_{q,p} = \int dx \varphi(x - qa) \varphi(x - pa) = \int dx \varphi(x) \varphi(x + (q - p)a) ,$$

$$U_{q,p} = V_0 a \sum_{n \neq p} \varphi((n - q)a) \varphi((n - p)a) , \quad (2.326)$$

$$(2.327)$$

and $s_{q,p} = s_{q-p} = s_{p-q}$, $U_{q,p} = U_{q-p} = U_{p-q}$. One finds

$$s_0 = 1 ,$$

$$s_1 = s_{-1} = (1 + \kappa a) e^{-\kappa a} ,$$

$$s_2 = s_{-2} = (1 + 2\kappa a) e^{-2\kappa a} . \quad (2.328)$$

It is found that s_1 and s_2 are proportional to $e^{-\kappa a}$ and $e^{-2\kappa a}$. For U_{p-q} or U_{q-p} one finds

$$U_0 = 2\varphi^2(a) V_0 a ,$$

$$U_1 = U_{-1} = \varphi(0) \varphi(a) V_0 a ,$$

$$U_2 = U_{-2} = (\varphi(0) \varphi(2a) + \varphi^2(a)) V_0 a . \quad (2.329)$$

Only the terms of the type $\varphi^2(a)$, $\varphi(0) \varphi(2a)$, $\varphi(0) \varphi(a)$ are taken into account in U_{p-q} or U_{q-p} to ensure the overlap integral up to second order. In the case of an infinite crystal, sums always run from $-\infty$ to $+\infty$. Considering the overlaps up to second order, the set of equations becomes

$$(\varepsilon_0 - \varepsilon + 2V_0 a \varphi^2(a)) \phi_q + [(\varepsilon_0 - \varepsilon) s_1 + V_0 a \varphi(0) \varphi(a)] (\phi_{q+1} + \phi_{q-1})$$

$$+ [(\varepsilon_0 - \varepsilon) s_2 + V_0 a (\varphi(0) \varphi(2a) + \varphi^2(a))] (\phi_{q+2} + \phi_{q-2}) = 0 . \quad (2.330)$$

After setting $\phi_q = e^{iqka}$, one finds for the energy band

$$\varepsilon = \varepsilon_0 + 2V_0 a \frac{\varphi^2(a) + \varphi(0) \varphi(a) \cos ka + (\varphi(0) \varphi(2a) + \varphi^2(a)) \cos(2ka)}{1 + 2s_1 \cos ka + 2s_2 \cos(2ka)} . \quad (2.331)$$

In the case of a crystal with finite size, $x_n = na$ with $n = 1, \dots, N$, and the potential given by $V(x) = V_0 a \sum_{n=1}^N \delta(x - na)$, (2.324) becomes:

$$\psi(x) = \sum_{p=1}^N \phi_p \varphi(x - pa) ,$$

while (2.325) becomes:

$$\sum_{p=1}^N \phi_p \left\{ (\varepsilon_0 - \varepsilon) s_{q,p} + \sum_{n \neq p, n=1}^N V_0 a \varphi((n-q)a) \varphi((n-p)a) \right\} = 0 , \quad (2.332)$$

with $1 \leq q \leq N$. The set of equation has now to be limited to first-neighbor overlaps, the inclusion of second neighbors resulting indeed quite complicated. From the set of equations (2.332) one thus obtains:

$$(\varepsilon_0 - \varepsilon) \phi_1 + [(\varepsilon_0 - \varepsilon) s_1 + V_0 a \varphi(0) \varphi(a)] \phi_2 = 0, \quad q = 1 , \quad (2.333)$$

and

$$(\varepsilon_0 - \varepsilon) \phi_q + [(\varepsilon_0 - \varepsilon) s_1 + V_0 a \varphi(0) \varphi(a)] (\phi_{q+1} + \phi_{q-1}) = 0, \quad q = 2, \dots, N-1 \quad (2.334)$$

and finally

$$(\varepsilon_0 - \varepsilon) \phi_N + [(\varepsilon_0 - \varepsilon) s_1 + V_0 a \varphi(0) \varphi(a)] \phi_{N-1} = 0, \quad q = N . \quad (2.335)$$

Expressions (2.334) are satisfied after setting $\phi_q = B \sin(q\sigma)$, if

$$\varepsilon = \varepsilon_0 + V_0 a \frac{2\varphi(0)\varphi(a) \cos \sigma}{1 + 2s_1 \cos \sigma} .$$

In this case (2.333) is automatically satisfied, while (2.335) does it if $\sigma = v\pi/(N+1)$ with v an integer.

2.4. Calculate the molecular orbitals of a 1D potential at two centers $V_0 a [\delta(x + \frac{d}{2}) + \delta(x - \frac{d}{2})]$.

Solution. The Schrödinger equation

$$-\frac{\hbar^2}{2m} \varphi''(x) + V_0 a \left[\delta\left(x + \frac{d}{2}\right) + \delta\left(x - \frac{d}{2}\right) \right] \varphi(x) = \varepsilon_0 \varphi(x) .$$

is solved writing the solutions in the regions $x < -d/2$, $-d/2 < x < d/2$, and $x > d/2$:

$$\varphi(x) = \begin{cases} A & e^{\kappa(x+\frac{d}{2})}, & x < -\frac{d}{2} \\ A_1 e^{-\kappa x} + A_2 e^{\kappa x}, & -\frac{d}{2} < x < \frac{d}{2} \\ B & e^{-\kappa(x-\frac{d}{2})}, & x > \frac{d}{2} \end{cases}$$

with $\kappa = \sqrt{2m\varepsilon_0}/\hbar$. Continuity conditions of the wavefunction and the jumps of the derivatives at $x = \pm d/2$ lead to the following equations:

$$\begin{aligned} A &= A_1 e^{\kappa \frac{d}{2}} + A_2 e^{-\kappa \frac{d}{2}}, \\ B &= A_1 e^{-\kappa \frac{d}{2}} + A_2 e^{\kappa \frac{d}{2}}, \end{aligned} \quad (2.336)$$

$$\begin{aligned} \frac{\kappa - \kappa_0}{\kappa} A &= -A_1 e^{\kappa \frac{d}{2}} + A_2 e^{-\kappa \frac{d}{2}}, \\ \frac{\kappa - \kappa_0}{\kappa} B &= A_1 e^{-\kappa \frac{d}{2}} - A_2 e^{\kappa \frac{d}{2}}, \end{aligned} \quad (2.337)$$

where $\kappa_0 = -2mV_0a/\hbar^2$. Substitution of (2.336) into (2.337) leads to a homogeneous set of two equations. The solutions are found through the equations

$$\left(2 - \frac{\kappa_0}{\kappa}\right) e^{\kappa \frac{d}{2}} = \frac{\kappa_0}{\kappa} e^{-\kappa \frac{d}{2}} \quad (2.338)$$

and

$$\left(2 - \frac{\kappa_0}{\kappa}\right) e^{\kappa \frac{d}{2}} = -\frac{\kappa_0}{\kappa} e^{-\kappa \frac{d}{2}}. \quad (2.339)$$

The first one requires that $2 - \kappa_0/\kappa > 0$, or else $\kappa/\kappa_0 > 1/2$, and the second one that $2 - \kappa_0/\kappa < 0$, or else $\kappa/\kappa_0 < 1/2$. In any event, one has

$$\begin{aligned} A_1 = A_2 &= \frac{A}{2 \cosh(\kappa \frac{d}{2})}, \quad A = B; \\ A_1 = -A_2 &= \frac{A}{2 \sinh(\kappa \frac{d}{2})}, \quad A = -B. \end{aligned} \quad (2.340)$$

Labeling as κ_1 and κ_2 the solutions of the two equations (2.338)-(2.339) with energies ε_1 and ε_2 respectively, their normalization coefficients A' and A'' are

$$\begin{aligned} A' &= \frac{\sqrt{\kappa_1}}{\left[1 + \tanh\left(\kappa_1 \frac{d}{2}\right) + \frac{\kappa_1 \frac{d}{2}}{\cosh^2\left(\kappa_1 \frac{d}{2}\right)}\right]^{\frac{1}{2}}}, \\ A'' &= \frac{\sqrt{\kappa_2}}{\left[1 + \coth\left(\kappa_2 \frac{d}{2}\right) - \frac{\kappa_2 \frac{d}{2}}{\sinh^2\left(\kappa_2 \frac{d}{2}\right)}\right]^{\frac{1}{2}}}. \end{aligned} \quad (2.341)$$

2.5. In the Plane Wave Method, the degeneracy of the unperturbed free-particle states is four and the matrix elements of the periodic potential satisfy the properties $V_{ij} = V_{ji}$ and $V_{ii} = V_0$. Calculate the energy when: a) $V_{12} = V_{13} = V_{24} = V_{34} = V_1$ and $V_{14} = V_{23} = 0$; b) $V_{12} = V_{13} = V_{24} = V_{34} = 0$ and $V_{14} = V_{23} = V_2$; c) all the matrix elements assume the same value V_0 .

Solution. The set of equations for the coefficients c_i with $i = 1, \dots, 4$ is

$$\begin{aligned} (\bar{E}_0 - E) c_1 + V_{12} c_2 + V_{13} c_3 + V_{14} c_4 &= 0, \\ V_{12} c_1 + (\bar{E}_0 - E) c_2 + V_{23} c_3 + V_{24} c_4 &= 0, \end{aligned}$$

$$\begin{aligned} V_{13}c_1 + V_{23}c_2 + (\bar{E}_0 - E)c_3 + V_{34}c_4 &= 0, \\ V_{14}c_1 + V_{24}c_2 + V_{34}c_3 + (\bar{E}_0 - E)c_4 &= 0, \end{aligned} \quad (2.342)$$

with $\bar{E}_0 = V_{00} + E_0$. The determinant of the coefficients matrix yields a fourth order polynomial.

In the case a), one sets $\bar{E}_0 - E = x$ and finds:

$$\begin{bmatrix} x & V_1 & V_1 & 0 \\ V_1 & x & 0 & V_1 \\ V_1 & 0 & x & V_1 \\ 0 & V_1 & V_1 & x \end{bmatrix} = 0,$$

with solutions

$$\begin{aligned} x^2 &= 0, \\ x^2 &= 4V_1^2. \end{aligned} \quad (2.343)$$

Notice that degeneracy is partly removed.

In the case b) one has:

$$\begin{bmatrix} x & 0 & 0 & V_2 \\ 0 & x & V_2 & 0 \\ 0 & V_2 & x & 0 \\ V_2 & 0 & 0 & x \end{bmatrix} = 0,$$

leading to

$$(x^2 - V_2^2)^2 = 0,$$

and telling that two paired solutions exist with $x = \pm |V_2|$.

In the case c) the determinant providing the eigenvalues is:

$$\begin{bmatrix} x & V_0 & V_0 & V_0 \\ V_0 & x & V_0 & V_0 \\ V_0 & V_0 & x & V_0 \\ V_0 & V_0 & V_0 & x \end{bmatrix} = 0,$$

with $x = E_0 + V_0 - E$. Subtracting the second row from the first, the third from the second, and the fourth from the third one finds:

$$\begin{bmatrix} x - V_0 & V_0 - x & 0 & 0 \\ 0 & x - V_0 & V_0 - x & 0 \\ 0 & 0 & x - V_0 & V_0 - x \\ V_0 & V_0 & V_0 & x \end{bmatrix} = 0.$$

The calculation yields three identical solutions $x = V_0$, or else $E = E_0$, and one solution equal to $x = -V_0$, or else $E = E_0 + 2V_0$.

2.6. Consider three generic crystals, with simple-cubic (SC), face-centered (FCC), and body-centered (BCC) structures. Starting from an s-orbital and considering only

first-neighbor overlaps, show that the band structures are given by the following expressions:

$$\begin{aligned}
 E(\mathbf{k}) &= E_0 + 2\gamma_1 [\cos(k_x a) + \cos(k_y a) + \cos(k_z a)]; \\
 E(\mathbf{k}) &= E_0 + 4\gamma_2 \left[\cos\left(\frac{k_x a}{2}\right) \cos\left(\frac{k_y a}{2}\right) + \cos\left(\frac{k_x a}{2}\right) \cos\left(\frac{k_z a}{2}\right) + \right. \\
 &\quad \left. \cos\left(\frac{k_y a}{2}\right) \cos\left(\frac{k_z a}{2}\right) \right]; \\
 E(\mathbf{k}) &= E_0 + 8\gamma_3 \cos\left(\frac{k_x a}{2}\right) \cos\left(\frac{k_y a}{2}\right) \cos\left(\frac{k_z a}{2}\right); .
 \end{aligned} \tag{2.344}$$

Solution. See the solution to the next Problem 2.7.

2.7. Under the conditions of Problem 2.6, calculate the effective masses at the Γ point in the BZ, and show that they are isotropic in all the three cases. Repeat the calculation at the BZ boundary along the directions $(1, 0, 0)$ and $(1, 1, 1)$.

Solution. For all the three cases, the overlap integrals with first-neighbors yield the same value. Thus, the six first-neighbors of simple cubic with coordinates $(\pm a, 0, 0)$, $(0, \pm a, 0)$, $(0, 0, \pm a)$ yield:

$$E(\mathbf{k}) = E_0 + \gamma_1 \left[e^{ik_x a} + e^{-ik_x a} + e^{ik_y a} + e^{-ik_y a} + e^{ik_z a} + e^{-ik_z a} \right],$$

resulting directly into the wanted expression, first line of (2.344).

The face-centered cubic lattice has twelve first neighbors, four in $x - y$ plane, four in $x - z$ plane, and four in $y - z$ plane. Using the same procedure of the previous case for the planes $x - y$, $x - z$ and $y - z$ respectively, the second line of (2.344) is obtained.

A body-centered cubic lattice has eight first-neighbors with coordinates $(\pm a, \pm a, \pm a)$. With the same procedure of the previous cases, the third line of (2.344) is obtained.

As to the effective masses, for the simple-cubic one has

$$M_{ij}^{-1} = \frac{1}{\hbar^2} \frac{\partial^2 E(\mathbf{k})}{\partial k_i \partial k_j} = -\frac{2\gamma_1 a^2}{\hbar^2} \delta_{ij} \cos(k_i a),$$

getting the special values $M_{ii}^{-1} = -2\gamma_1 a^2 / \hbar^2$ at the Γ point with $k = 0$, $M_{xx}^{-1} = 2\gamma_1 a^2 / \hbar^2$, $M_{yy}^{-1} = M_{zz}^{-1} = -2\gamma_1 a^2 / \hbar^2$ at the X point with $k \equiv (\pi/a, 0, 0)$, and $M_{ii}^{-1} = 2\gamma_1 a^2 / \hbar^2$ at the L point with $k \equiv (\pi/a, \pi/a, \pi/a)$.

For a face-centered cubic one has for M_{xx}^{-1} and M_{xy}^{-1} :

$$\begin{aligned}
 M_{xx}^{-1} &= -\frac{\gamma_2 a^2}{\hbar^2} \cos\left(\frac{k_x a}{2}\right) \left[\cos\left(\frac{k_y a}{2}\right) + \cos\left(\frac{k_z a}{2}\right) \right] \\
 M_{xy}^{-1} &= \frac{\gamma_2 a^2}{\hbar^2} \sin\left(\frac{k_x a}{2}\right) \sin\left(\frac{k_y a}{2}\right) = M_{yx}^{-1}.
 \end{aligned} \tag{2.345}$$

The other quantities are obtained by a convenient interchange of x , y and z symbols. Special values are at $\Gamma \equiv (0,0,0)$ with isotropic mass $M = -\hbar^2/(2\gamma_2 a^2)$, since $M_{ij}^{-1} = 0$ for $i \neq j$. At point $X \equiv (2\pi/a, 0, 0)$, one finds $M_{xx}^{-1} = 2\gamma_2 a^2/\hbar^2$, $M_{yy}^{-1} = M_{zz}^{-1} = 0$, and $M_{ij}^{-1} = 0$ for $i \neq j$. At point $L \equiv (2\pi/a, 2\pi/a, 2\pi/a)$, one finds $M_{xx}^{-1} = M_{yy}^{-1} = M_{zz}^{-1} = -2\gamma_2 a^2/\hbar^2$, and $M_{ij}^{-1} = 0$ for $i \neq j$.

Finally, for a body-centered cubic one has:

$$\begin{aligned} M_{xx}^{-1} &= -\frac{2\gamma_3 a^2}{\hbar^2} \cos\left(\frac{k_x a}{2}\right) \cos\left(\frac{k_y a}{2}\right) \cos\left(\frac{k_z a}{2}\right) = M_{yy}^{-1} = M_{zz}^{-1} \\ M_{xy}^{-1} &= \frac{2\gamma_3 a^2}{\hbar^2} \sin\left(\frac{k_x a}{2}\right) \sin\left(\frac{k_y a}{2}\right) \cos\left(\frac{k_z a}{2}\right) = M_{yx}^{-1}. \end{aligned} \quad (2.346)$$

The other quantities are obtained by a convenient interchange of x , y and z symbols. The diagonal values of the masses result to be equal and are: $M^{-1} = -2\gamma_3 a^2/\hbar^2$ at point $\Gamma \equiv (0, 0, 0)$, $M^{-1} = 2\gamma_3 a^2/\hbar^2$ at point $H \equiv (2\pi/a, 0, 0)$, and $M^{-1} = 2\gamma_3 a^2/\hbar^2$ at point $P \equiv (2\pi/a, 2\pi/a, 2\pi/a)$. Off-diagonal values M_{ij}^{-1} at all the three points Γ , H and P are zero.

2.8. Determine the surface states of a system with N particles described by the model equations

$$\begin{aligned} (E'_0 - E)\phi_1 + \gamma\phi_2 &= 0, \\ (E_0 - E)\phi_n + \gamma(\phi_{n+1} + \phi_{n-1}) &= 0 \text{ for } n = 2, \dots, N-1, \\ (E'_0 - E)\phi_N + \gamma\phi_{N-1} &= 0. \end{aligned} \quad (2.347)$$

Solution. Choosing the coefficients ϕ_n in the form

$$\phi_n = A e^{inx} e^{-ny} + B e^{i(N-n)x} e^{-n(N-n)y},$$

with x and y (> 0) being the real and imaginary parts of ka respectively. By substitution in the second of the three equations, it is found that these equations are satisfied for any $n = 2, \dots, N-1$ if

$$E_0 - E + \gamma \left\{ e^{(-ix+y)} + e^{(ix-y)} \right\} = 0.$$

The imaginary part vanishes when $\sin(x) = 0$, that is for $x = p\pi$ with p an integer. Setting $E_1 = E_0 - E$, one finds:

$$\cosh y = -(-1)^p \frac{E_1}{2\gamma}. \quad (2.348)$$

The value of p is fixed so that the right-hand-side be positive. It must also be $|E_1/(2\gamma)| > 1$. The first and third equations are fulfilled, provided that

$$A e^{(ix-y)} \left[E'_0 - E + \gamma e^{(ix-y)} \right] + B e^{(N-1)(ix-y)} \left[E'_0 - E + \gamma e^{(-ix+y)} \right] = 0,$$

$$Ae^{N(ix-y)} \left[E'_0 - E + \gamma e^{(-ix+y)} \right] + B \left[E'_0 - E + \gamma e^{(ix-y)} \right] = 0. \quad (2.349)$$

The determinant of the above coefficients matrix vanishes if

$$\left[E'_0 - E + \gamma e^{(ix-y)} \right]^2 - e^{2(N-1)(ix-y)} \left[E'_0 - E + \gamma e^{(-ix+y)} \right]^2 = 0.$$

Considering the possible values of x and setting $E_2 = E'_0 - E$, one then obtains:

$$\begin{aligned} & E_2^2 + 2E_2\gamma(-1)^p e^{-y} + \gamma^2 e^{-2y} \\ & - e^{-2(N-1)y} \left[E_2^2 + 2E_2\gamma(-1)^p e^y + \gamma^2 e^{2y} \right] = 0. \end{aligned} \quad (2.350)$$

This is a quadratic equation in the variable E_2 with solutions

$$\frac{E_2}{\gamma} = \frac{-(-1)^p \left[e^{-y} - e^{(-2(N-1)y+y)} \right] \pm e^{-(N-1)y} (e^y - e^{-y})}{1 - e^{-2(N-1)y}}.$$

Substitution of the value of y from (2.348) in the above equation, leads to an equation for the unknown E to be numerically solved. The two resulting solutions correspond to two surface states localized at the two very ends of the chain. The $N \rightarrow \infty$ limit of the last equation has to correspond to (2.34) in the text.

2.9. Show that the energy of a surface state E_s like that in (2.35), is larger than $E_M = E_0 + 2|\gamma|$ when $E'_0 > E_0$, and smaller than $E_m = E_0 - 2|\gamma|$ when $E'_0 < E_0$.

Solution. Since

$$E_s = E'_0 + \frac{\gamma^2}{E'_0 - E_0}$$

it is easy to see that $E_s - E_M > 0$ and $E_s - E_m < 0$

2.10. Show that the function

$$\tan[(N+1)ka] = \frac{\sin(ka)}{\cos(ka) - \frac{\gamma}{E'_0 - E_0}}$$

has N real solutions for $0 < ka < \pi$ and $|\gamma/(E'_0 - E_0)| \geq 1$ and $N-1$ real solutions for $0 < ka < \pi$ and $|\gamma/(E'_0 - E_0)| < 1$.

Setting $ka = x$, the right-hand side function has vertical asymptotes when $x = (2n+1)\pi/[2(N+1)]$ and has a zero at $x = n\pi/(N+1)$ for $n = 0, 1, \dots$. In the range $0 < x < \pi$, the largest coordinate asymptote X is given by the relation $X = (2N' + 1)\pi/[2(N+1)] < \pi$. The largest integer N' that satisfies the relation is $N' = N-1$, the largest coordinate of the asymptote in $0 < x < \pi$ is $X_m = (N-1/2)\pi/(N+1)$, and the largest coordinate of the zero is

$$a_m = \frac{N - \frac{1}{2}}{N+1} \pi + \frac{\pi}{N+1} = \frac{\pi}{N+1} \left(N + \frac{1}{2} \right) < \pi.$$

The values of n that satisfy the given conditions are N ($n = 0, 1, \dots, N - 1$). If $\xi = |\gamma/(E'_0 - E_0)| > 1$, the function at the right-hand side has a well defined sign and it vanishes at $x = 0, \pi$. For this reason there are N crossings between the functions at the left and right hand sides of the given equation.

If $\xi = |\gamma/(E'_0 - E_0)| < 1$, similar considerations lead to the observation that the function to the right-hand side has one vertical asymptote: the number of real solutions is thus reduced to $N - 1$.

Problems without solutions

2.11. Consider the one-dimensional band structure

$$E(k) = E_0 + 2\gamma_1 \cos(ka) + 2\gamma_2 \cos(2ka) .$$

Calculate the effective mass close to the points $k = 0$ and $k = \pi/a$, and determine the condition that must be satisfied to have a critical point inside the Brillouin zone.

2.12. Calculate the effective mass at the critical point inside the Brillouin zone determined in Problem 2.11.

2.13. Show that each among (2.338) and (2.339) admits only one solution.

2.14. Assuming $\gamma < 0$, draw from (2.11) the energy levels $E(k) - E_0$ for nanostructures with $N = 3$ and $N = 4$ atoms.

2.15. Calculate the energy levels for a system with N particles when all interaction matrix elements are equal.

2.16. Consider the band structure

$$\varepsilon(k) = \varepsilon_0 + 2\gamma_1 \cos(ka) + 2\gamma_2 \sin^2(ka) .$$

Show that it is related to a system with interactions up to second neighbors.

2.17. Systems with two atomic species have been studied in Sec. 2.2.6 for a finite chain with given boundary conditions and two values for the on-site energy, and in Sec. 2.3.2 for a periodic chain with different interaction parameters among first neighbors. The band structure is given in the former case by (2.44) and by (2.81) in the latter. Elaborate similarities and differences between the two problems.

2.18. Consider the band structure for a square lattice

$$E(\mathbf{k}) = E_0 + 2\gamma(\cos(k_x a) + \cos(k_y a)) ,$$

with $\gamma < 0$. Find the critical point and the effective masses in the Brillouin zone.

2.19. Express the electronic part of the Hamiltonian for the hydrogen molecule in the Born-Oppenheimer approximation.

2.20. Calculate the Langevin function (2.207).

References

1. L. Bloomfield: *How Everything Works. Making physics out of ordinary.* John Wiley and Sons, USA (2008)
2. S. M. Gaponenko: *Optical properties of semiconductors nanocrystals.* 1st ed. Cambridge University Press, Cambridge (1998)
3. N.W. Ashcroft and N. D. Mermin: *Solid-State Physics.* 2nd ed. Holt-Saunders, Tokyo (1976)
4. F. Bassani and U.M. Grassano: *Fisica dello Stato Solido.* 1st ed. Bollati Boringhieri, Torino (2000)
5. J. M. Ziman: *Principles of the theory of solids.* 2nd ed. Cambridge University Press, Cambridge (1964)
6. J.C. Slater: *Quantum theory of Matter.* 5th ed. McGraw-Hill Book Company, New York (1985)
7. N. Nilius, T. M. Wallis, W. Ho, *Science* **297**, 1853-1856 (2002)
8. G. Grosso and G. Pastori Parravicini: *Solid State Physics.* 1st ed. Academic Press, New York (2000)
9. P. Lauffer, K. V. Emtsev, R. Graupner, Th. Saylor, L. Ley, S. A. Reshanov, and H. B. Weber, *Phys. Rev. B* **77**, 155426-155435 (2008)
10. J.C. Charlier, X. Blase, S. Roche, *Rev. Mod. Phys.* **79**, 677-732 (2007)
11. E. Kaxiras: *Atomic and Electronic Structure of Solids.* 1st ed. Cambridge Univ. Press (2003)
12. G. D. Mahan: *Many-Particle Physics.* 3rd ed. Kluwer Academic, New York (2000)
13. J.C. Slater, *Phys. Rev.* **87**, 807-835 (1952)
14. I. Bloch, *Nature* **453**, 1016 (2008)
15. E.Y. Tsimbal, in *Physics 927*, available at http://physics.unl.edu/tsymbal/teaching/SSP-927/Section%2011_Methods_for_calculating_band_structure.pdf G.A. Bordick, *Phys. Rev.* **129**, 138 (1963)
16. Lin Wang Wang and A. Zunger, *J. Phys. Chem.* **98**, 2158-2165 (1994)
17. V. P. Gribkovskii, L. G. Zimin, S. V. Gaponenko, I. E. Malinovskii, P. I. Kuznetsov, and G. Yakushcheva, *Physica Status Solidi (b)* **158**, 359-366 (1990)
18. H. Hofmeister, F. Huysken, and B. Kohn, *Eur. Phys. Jour. D* **9**, 137-140 (1999)
19. H. Ichinose, *Sci. Tech. Adv. Mat.* **1**, 11-20 (2000)
20. Xiaogang Peng, L. Manna, Weldong Yang, J. Wickam, E. Scher, A. Kadavanich, and A.P. Allvisatos, *Nature* **404**, 59-62 (2000)
21. G. Cantele, D. Ninno, and G. Iadonisi, *J. Phys.: Condens. Matter* **12**, 9019-9026 (2000)
22. Y. Kayanuma, *Solid State Comm.* **59**, 405-408 (1986)
23. Y. Kayanuma, *Phys. Rev. B* **38**, 9797-9805 (1988)
24. H. Leon, J. L. Marin, and R. Riera, *Physica E* **27**, 385-396 (2005)
25. C. Kittel: *Introduction to Solid State Physics.* 6th ed. Wiley, New York (1986)
26. P. Weiss and R. Forrer, *Ann. Physik* **5**, 153 (1926)
27. R.N. Sinclair and B.N. Brockhouse, *Phys. Rev.* **120**, 1638 (1960)
28. M. H. Anderson, J. R. Ensher, M. R. Matthews, C. E. Wieman, and E. A. Cornell, *Science* **269**, 198 (1995); K. B. Davis, M.-O. Mewes, M. R. Andrews, N. J. van Druten, D. S. Durfee, D. M. Kurn, and W. Ketterle, *Phys. Rev. Lett.* **75**, 3969 (1995)
29. H. K. Onnes, *Comm. Phys. Lab. Univ. Leiden* **12** (1911)
30. J. Bardeen, L. N. Cooper, and J. R. Schrieffer, *Phys. Rev.* **106**, 162 (1957)
31. J. G. Bednorz and K. A. Müller, *Z. Physik, B* **64**, 189 (1986)

32. M. J. Holland, S. Kokkelmans, M. L. Chiofalo, R. Walser, *Phys. Rev. Lett.* **87**, 120406 (2001)
33. M. Greiner, C. A. Regal, and D. S. Jin, *Nature* **426**, 537 (2003)
34. Y. J. Uemura *et al.*, *Nature* **352**, 605 (1991)
35. H. K. Onnes, *Akad. Wetenschappen (Amsterdam)* **14**, 113 (1911)
36. D. M. Ginsberg and M. Tinkham, *Phys. Rev.* **118**, 990 (1960)
37. V. Phillips, *Phys. Rev.* **134**, A 385 (1964)
38. P. Townsend and J. Sutton, *Phys. Rev.* **128**, 591 (1962)
39. D. L. Decker, D. E. Mapother, and R. W. Shaw, *Phys. Rev.* **112**, 1888 (1958)
40. I. M. Khalatnikov: *An Introduction to the Theory of Superfluidity*. Benjamin, INC., New York (1965)
41. For a nice pedagogical treatment see parts of G. Baym: in *Mathematical Methods in Solid State and Superfluid Theory*. R.C. Clark and G. H. Derrick eds. Oliver and Boyd, Edinburgh (1969)
42. For a review, see e.g. *Models and Phenomenology for Conventional and High-Temperature Superconductivity*. Proc. of Intl. School of Physics "Enrico Fermi", Course CXXXVI. G. Iadonisi, J. R. Schrieffer, and M. L. Chiofalo eds. IOS Press, Amsterdam (1998)
43. D. M. Eagles, *Phys. Rev.* **186**, 456 (1969)
44. A. J. Leggett: *Modern Trends in the Theory of Condensed Matter*. A. Pekalski and R. Przystawa eds. Lecture Notes in Physics **115**, 13. Springer-Verlag, Berlin (1980)
45. P. Y. Yu and M. Cardona: *Fundamentals of Semiconductors*. Springer, Berlin (1996)
46. Division of Electron Microscopy and Crystal Chemistry Institute for Chemical Research, available at <http://eels.kuicr.kyoto-u.ac.jp/eels.en.html>
47. F. Reinert and S. Hüfner, *New J. Phys.* **7**, 97 (2005)
48. S. Hüfner: *Photoelectron Spectroscopy. Principles and Applications* 3rd edn. Berlin, Springer (2003)
49. D. Y. Joh, L. H. Herman, San-Yong Ju, J. Kinder, M. A. Segal, J. N. Johnson, G. K. L. Chan, and J. Park, *Nano Lett.* **11**, 1 (2011)
50. Kazunari Matsuda (2011). *Exciton Dephasing in a Single Carbon Nanotube Studied by Photoluminescence Spectroscopy*, *Electronic Properties of Carbon Nanotubes*, Prof. Jose Mauricio Marulanda (Ed.). ISBN: 978-953-307-499-3, InTech, DOI: 10.5772/19652. Available from: <http://www.intechopen.com/books/electronic-properties-of-carbon-nanotubes/exciton-dephasing-in-a-single-carbon-nanotube-studied-by-photoluminescence-spectroscopy>
51. See e.g. P. Nozières and D. Pines: *The Theory of Quantum Liquids - Superfluid Bose Liquids*. Addison-Wesley, New York (1990)
52. M.H. Anderson, J.R. Ensher, M.R. Matthews, C.E. Wieman, and E.A. Cornell, *Science* **269**, 198 (1995)
53. K.B. Davis, M.O. Mewes, N.J. van Druten, D.S. Durfee, D.M. Kurn, and W. Ketterle, *Phys. Rev. Lett.* **75**, 3969 (1995)
54. For a review, see e.g. L.P. Pitaevskii and S. Stringari: *Bose-Einstein Condensation*. Oxford University Press, Oxford (2003)

Chapter 3

Elements of continuum mechanics and lattice vibrations

Abstract Two related issues are here discussed: a) deformations of continuous bodies responding to applied external forces, their dynamics within a classical approach and applications of related concepts to both macroscopic and nanoscopic systems; b) classical dynamics of systems made of lattice points with finite and infinite size, with varying boundary conditions. Issue b) provides a microscopical framework to understand the macroscopic phenomena discussed in a). Classical dynamics represented in terms of normal modes is shown to be an useful framework to pass to the quantum description, and introduce the phonons as vibrational quanta. Applications to specific heats of insulators are illustrated, including the effects of anharmonic vibrations. Finally, experimental methods and techniques are discussed, related to the determination of vibrational spectra in solids.

3.1 Introduction

Did you ever ask yourself how it happens that different materials have different reactions to deformation and stresses, some of them being elastic, other plastic, or brittle or hard. Or else, how it occurs that different materials sound in different ways when hit, or how musical instruments or else quartz clocks work? How it happens that materials expand or shrink when subjected to temperature changes? Eventually, are all these questions connected and if they do, how can the answers be hardwired together in the same theoretical framework? The content of this chapter may open up useful answers [1].

Whenever a continuous or a microscopic deformation is applied to a material, an energy cost must be paid to produce such a deformation. Along these lines, the study of the cohesive energy performed in Chap. 1 is a property of the ground state, corresponding to atoms at their equilibrium positions. This Chapter deals with the atomic displacements with respect to the *perfect* lattice configuration, and investigates the manifestations of their occurrence.

The first part of the present Chapter is devoted to studying deformations of a non particle-like body, bridging the content of this Chapter with the discussion on mechanical properties and crystal structure at the end of Chap. 1. In this approach, a body is viewed as a continuum where the distribution of matter is described by a continuous function, the density $\rho(\mathbf{r})$, while the body is subjected to a stress. The presentation closely follows the pedagogical notes of Fausto Fumi [2]. The relationships between deformation and stress are considered in the elastic limit, that is when the original form is restored (deformation disappears) once the stress vanishes. It is then possible to express the system behavior by means of linear constitutive equations relating deformation and stress. The equation coefficients depend on the material nature and symmetry properties. The laws of classical dynamics for the finite body elements allows to classify the deformations that propagate in the medium. Selected applications are proposed, that can be in a few cases extended to nanosystems. Piezoelectricity and piezomagnetism are introduced at this stage of the textbook, as phenomena where the elastic material response is related to electric or magnetic properties instead than to mechanical stresses.

The theory of elastic deformations is a classical theory, basically providing the energy that is required to induce small deformations in a solid and the dynamics that results from those deformations, in the limit of wavelengths much longer than typical microscopic distances. On the other hand, the study of lattice vibrations via the quantum description and the concept of phonons, extends the study of deformations and their dynamics also to the case of wavelengths that are comparable to the typical atomic distances. Therefore, in a second part of the Chapter, the oscillations of atoms or molecules around the lattice equilibrium positions are studied within a classical dynamics framework. In the case of nanosystems, the solutions are found to depend on boundary conditions, the latter becoming immaterial in the limit of macroscopic systems. In the case of macroscopic systems, it is then found that special solutions of the molecular approach exist in the long-wavelength limit, that coincide with those obtained for bodies with continuous matter distributions.

Chap. 2 has introduced the Born-Oppenheimer approximation, where ions wavefunctions satisfy a Schrödinger equation with potential $\varepsilon(\mathbf{R})$ depending on ionic coordinates \mathbf{R} , that is given by the electron Hamiltonian eigenvalue. There, it is also shown that vibrational energies are much smaller than electronic ones. Starting from a minimum of $\varepsilon(\mathbf{R})$, it is then shown that a second-order Taylor expansion in ion coordinates, that defines the harmonic approximation, maps the problem of ionic motion into that of independent quantum harmonic oscillators, named phonons. Phonon frequencies result to coincide with those calculated by classical equations. The number of oscillators identifies the number of the system degrees of freedom.

The final part of the Chapter is devoted to studying manifestations of the physics of lattice vibrations, such as the specific heat of solids, and of anharmonic effects as in selected thermal properties. Eventually the experimental techniques are illustrated, that are aimed to determine the crystal vibrational spectra.

3.2 Macroscopic picture: deformations and stresses in continuous media

The introduction of a full set of parameters that describe the deformation originated by stresses in continuous media is the subject of detailed treatises in the framework of mathematical physics and engineering. This section introduces the main concepts in a minimal manner, in order to deal with an easily tractable mathematical representation while providing essential tools [2, 3].

Consider a homogeneous material, that is with constant density ρ , on which a deformation occurs as a consequence of a stress. While the definitions of stress and strain have been introduced in Sec. 1.8, a representation of uniform deformations is now needed.

Definition

A deformation is uniform when the orthogonal set of cartesian axes with respect to which the body deformation is represented, is transformed into a set of non-orthogonal axes as a consequence of deformation.

The original versors $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ of the orthogonal frame become vectors of non unitary length \mathbf{x}', \mathbf{y}' , and \mathbf{z}' . The general transformation equations are:

$$\begin{aligned}\mathbf{x}' &= (1 + \varepsilon_{xx})\hat{\mathbf{x}} + \varepsilon_{xy}\hat{\mathbf{y}} + \varepsilon_{xz}\hat{\mathbf{z}}, \\ \mathbf{y}' &= \varepsilon_{yx}\hat{\mathbf{x}} + (1 + \varepsilon_{yy})\hat{\mathbf{y}} + \varepsilon_{yz}\hat{\mathbf{z}}, \\ \mathbf{z}' &= \varepsilon_{zx}\hat{\mathbf{x}} + \varepsilon_{zy}\hat{\mathbf{y}} + (1 + \varepsilon_{zz})\hat{\mathbf{z}},\end{aligned}\tag{3.1}$$

where $(\delta_{i,j} + \varepsilon_{i,j})$ are the coefficients relating the old to the new axes. The linear transformation among versors $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ and vectors $\mathbf{x}', \mathbf{y}', \mathbf{z}'$ becomes the identical transformation ($\mathbf{x}' = \hat{\mathbf{x}}, \mathbf{y}' = \hat{\mathbf{y}}, \mathbf{z}' = \hat{\mathbf{z}}$) whenever $\varepsilon_{ij} = 0$, with $i, j \equiv x, y, z$. Each body element is then subjected to a displacement that can be represented as:

$$\begin{aligned}\mathbf{s}(\mathbf{r}) &= \mathbf{r}' - \mathbf{r} = x(\mathbf{x}' - \hat{\mathbf{x}}) + y(\mathbf{y}' - \hat{\mathbf{y}}) + z(\mathbf{z}' - \hat{\mathbf{z}}) \\ &= (x\varepsilon_{xx} + y\varepsilon_{yx} + z\varepsilon_{zx})\hat{\mathbf{x}} + (x\varepsilon_{xy} + y\varepsilon_{yy} + z\varepsilon_{zy})\hat{\mathbf{y}} \\ &\quad + (x\varepsilon_{xz} + y\varepsilon_{yz} + z\varepsilon_{zz})\hat{\mathbf{z}}.\end{aligned}\tag{3.2}$$

The meaning of the coefficients ε_{ij} is established after calculating the projections $\mathbf{x}' \cdot \mathbf{x}'$ and $\mathbf{x}' \cdot \mathbf{y}'$. In the former case one has:

$$\mathbf{x}' \cdot \mathbf{x}' = (1 + \varepsilon_{xx})^2 + \varepsilon_{xy}^2 + \varepsilon_{xz}^2 \simeq 1 + 2\varepsilon_{xx}\tag{3.3}$$

under the assumption $|\varepsilon_{ij}| \ll 1$, which is valid for small deformations. One then finds that ε_{ii} represents either the stretching or the compression of i -th axis. In the latter case, one has:

$$\mathbf{x}' \cdot \mathbf{y}' = \varepsilon_{yx}(1 + \varepsilon_{xx}) + \varepsilon_{xy}(1 + \varepsilon_{yy}) + \varepsilon_{xz}\varepsilon_{yz} \simeq \varepsilon_{yx} + \varepsilon_{xy}.\tag{3.4}$$

Thus, $\varepsilon_{ji} + \varepsilon_{ij}$ with $i \neq j$ provides the cosines of the angles that axes \mathbf{x}' , \mathbf{y}' , and \mathbf{z}' form between one another. Since the angles between the initial axes is $\pi/2$, the new angles $\pi/2 - \alpha$ are such that $\cos(\pi/2 - \alpha) = \sin \alpha \simeq \alpha = \varepsilon_{yx} + \varepsilon_{xy}$. The six quantities ε_{ii} with $i = 1, 2, 3$, and $\varepsilon_{ji} + \varepsilon_{ij}$ with $i \neq j$, are sufficient to describe the deformation.

A non-uniform deformation $\mathbf{s}(\mathbf{r})$ instead, can be represented as:

$$\mathbf{s}(\mathbf{r}) = u(\mathbf{r})\hat{\mathbf{x}} + v(\mathbf{r})\hat{\mathbf{y}} + w(\mathbf{r})\hat{\mathbf{z}}. \quad (3.5)$$

After expanding $u(\mathbf{r})$ in Taylor series to first order in the vicinity of $\mathbf{r} = 0$, one finds:

$$u(\mathbf{r}) = u(0) + x\partial_x u(\mathbf{r})|_{\mathbf{r}=0} + y\partial_y u(\mathbf{r})|_{\mathbf{r}=0} + z\partial_z u(\mathbf{r})|_{\mathbf{r}=0},$$

where the definitions $\partial_i u(\mathbf{r}) \equiv \partial u(\mathbf{r})/\partial i$, $i = x, y, z$ have been adopted. In similar manner for $v(\mathbf{r})$ and $w(\mathbf{r})$, one obtains:

$$\begin{aligned} \mathbf{s}(\mathbf{r}) = & u(0)\hat{\mathbf{x}} + v(0)\hat{\mathbf{y}} + w(0)\hat{\mathbf{z}} + (x\partial_x u + y\partial_y u + z\partial_z u)\hat{\mathbf{x}} \\ & + (x\partial_x v + y\partial_y v + z\partial_z v)\hat{\mathbf{y}} + (x\partial_x w + y\partial_y w + z\partial_z w)\hat{\mathbf{z}}. \end{aligned} \quad (3.6)$$

Equation (3.6) states that on a locale scale non-uniform deformations can be expressed in terms of uniform deformations. By comparison with (3.2), the following correspondence table can be constructed:

$$\begin{array}{lll} \varepsilon_{xx} = \partial_x u, & \varepsilon_{yx} = \partial_y u, & \varepsilon_{zx} = \partial_z u, \\ \varepsilon_{xy} = \partial_x v, & \varepsilon_{yy} = \partial_y v, & \varepsilon_{zy} = \partial_z v, \\ \varepsilon_{xz} = \partial_x w, & \varepsilon_{yz} = \partial_y w, & \varepsilon_{zz} = \partial_z w. \end{array} \quad (3.7)$$

It is evident that the relevant parameters here are all the possible scalar products between \mathbf{x}' , \mathbf{y}' , \mathbf{z}' . From (3.3), (3.4) and (3.7), six parameters result to locally describe a non-uniform deformation, that are:

$$\begin{array}{lll} e_{xx} = \partial_x u, & e_{yy} = \partial_y v, & e_{zz} = \partial_z w, \\ e_{xy} = \partial_y u + \partial_x v, & e_{yz} = \partial_z v + \partial_y w, & e_{zx} = \partial_x w + \partial_z u. \end{array} \quad (3.8)$$

The stresses are now to be represented. Consider a unit surface and a force acting on it. The latter can be decomposed into one component normal to the surface and two components tangential to it. Consider the special case of a cube with unit side: the force acting on a surface perpendicular to the x axis has components X_x , Y_x and Z_x . In similar manner, forces acting on surfaces perpendicular to the axes y and z are respectively X_y , Y_y , Z_y and X_z , Y_z , Z_z . Consider an elementary cube with infinitesimal sides dx , dy and dz as a rigid body. The elementary cube can be acted by external forces applied on any point of it, as in the case of electric or magnetic fields. It can as well be acted by internal contact forces that parts of the body, external to the elementary cube, exert through its delimiting surfaces. At equilibrium, both net forces and net torques are to be zero. Under this condition, that is with only internal contact forces acting on it, it must be

$$Y_z = Z_y, \quad Z_x = X_z, \quad X_y = Y_x. \quad (3.9)$$

Under non-equilibrium conditions, net forces and torques are non-vanishing, but one can show that the three conditions (3.9) are to be once again satisfied in order to prevent singular motions such as occurrence of vortexes. It follows that only six among the nine components of the forces are independent, as many as the deformation parameters. A compact notation can thus be used, that is:

$$\begin{aligned} e_1 &= e_{xx}, \quad e_2 = e_{yy}, \quad e_3 = e_{zz}, \\ e_4 &= e_{yz}, \quad e_5 = e_{zx}, \quad e_6 = e_{xy}, \\ T_1 &= X_x, \quad T_2 = Y_y, \quad T_3 = Z_z, \\ T_4 &= Y_z, \quad T_5 = Z_x, \quad T_6 = X_y. \end{aligned} \quad (3.10)$$

The connection between stresses and deformations is expressed by the constitutive relations, determining the stresses needed to originate given deformations, and viceversa.

Definition

For small stresses and deformations, their connection is represented by the linear equations:

$$e_i = \sum_{j=1}^6 s_{ij} T_j, \quad (3.11)$$

defining the *compliance* coefficients s_{ij} .

Definition

The inverse relations

$$T_i = \sum_{j=1}^6 c_{ij} e_j \quad (3.12)$$

define instead the *stiffness* coefficients c_{ij} .

Equations (3.11)-(3.12) are the constitutive relations, representing the connections between stresses and deformations that are built-in, constitutive indeed, on the way a material responds to stress. One assumes that the coefficients determinant be non vanishing, so that vanishing stresses correspond to vanishing deformations and viceversa.

Definition

Conditions can be given so that once the stress is removed, body's shape is restored to its original form, namely the deformation vanishes. This is the elastic limit.

Purely thermodynamic considerations lead to conclusion that the coefficients c_{ij} and s_{ij} are symmetric, that is $c_{ij} = c_{ji}$ and $s_{ij} = s_{ji}$, as elaborated in Problem 3.1. Thus, the number of independent elastic constants is at most 21, and not 36. The symmetries of the point group further reduce the number of non-vanishing coefficients: in the case of cubic symmetry for example, they are reduced to the following: $c_{ii} = c_{11}$ for $i = 1, 2, 3$, $c_{ii} = c_{44}$ for $i = 4, 5, 6$, and $c_{12} = c_{13} = c_{23}$, as shown in Problem 3.2.

The constitutive relations are an essential tool to calculate the response of the system to small mechanical perturbations, that is within a linear approximation. The following procedure can be typically used:

Procedure

- Step 1.** Begin with the constitutive relations (3.12).
 - Step 2.** Apply the external stress T_i along the appropriate directions.
 - Step 3.** Use and include all possible symmetry relations between the stiffness coefficients c_{ij} .
 - Step 4.** Define the wanted response, that is the wanted deformation in terms of the wanted stress.
 - Step 5.** Solve the set of linear equations so to get the deformation-stress equation, which identifies the wanted response coefficient.
-

Examples

In the following, examples are given where this procedure is applied to determine rigidity and related compressibility, shear coefficient and related shape elasticity. The latter represent, respectively, the response of the system to a stretching/compression or to a shear stress that does not change the system volume.

3.2.1 Rigidity, compressibility, and shape elasticity

Consider a material with cubic symmetry subjected to forces X_x and $-X_x$ acting on the two orthogonal faces to x -axis. From (3.12) one obtains:

$$\begin{aligned} X_x &= c_{11}e_1 + c_{12}(e_2 + e_3), & 0 &= Y_z = c_{44}e_{yz}, \\ 0 &= c_{12}e_1 + c_{11}e_2 + c_{12}e_3, & 0 &= Z_x = c_{44}e_{zx}, \\ 0 &= c_{12}e_1 + c_{12}e_2 + c_{11}e_3, & 0 &= X_y = c_{44}e_{xy}. \end{aligned} \quad (3.13)$$

The expressions on the right column lead to the result $e_{yz} = e_{zx} = e_{xy} = 0$, implying that angles between the primed axes \mathbf{x}' , \mathbf{y}' and \mathbf{z}' are once more $\pi/2$. Second and third equations on the left column instead lead to the results $e_2 = e_3$ and

$$\frac{e_2}{e_1} = -\frac{c_{12}}{c_{11} + c_{12}}. \quad (3.14)$$

Finally, the first equation on the left column implies:

$$X_x = \left(c_{11} - \frac{2c_{12}^2}{c_{11} + c_{12}} \right) e_1. \quad (3.15)$$

Rigidity and Young modulus. The expressions above define two relevant parameters as follows.

Definition

The quantity $\sigma = c_{12}/(c_{11} + c_{12})$ in (3.14) expresses the stretching of unit vector \hat{y} relative to \hat{x} . It is called Poisson coefficient, its experimental value usually being smaller than 0.5.

Consider now a parallelepiped bar with length ℓ and section A subjected to a normal force F . One then has $X_x = F/A$. Looking at (3.15) and given that e_1 represents the stretching of a unit-length segment, one obtains

$$\begin{aligned} \frac{\Delta\ell}{\ell} &= e_1 = \frac{1}{G} \frac{F}{A}, \\ G &\equiv c_{11} - \frac{2c_{12}^2}{c_{11} + c_{12}}. \end{aligned} \quad (3.16)$$

Definition

The quantity G^{-1} expresses the proportionality coefficient between the bar length % variation $\Delta\ell/\ell$ and the force per unit area F/A . G is called Young modulus and is a measure of material rigidity.

Concept

A stress per unit area F/A acted orthogonally on a side of the material results into a deformation $\Delta\ell$. In the limit of small acting stresses, the relative deformation $\Delta\ell/\ell$ has a linear-law dependence on F/A like that displayed in Fig. 1.41, with a G^{-1} proportionality coefficient and a vanishing intercept. This coefficient measures the linear material response to a compression. Thus: a) the larger the material rigidity, the smaller is the deformation at given stress, and b) if the stress is removed, the deformation vanishes. This is the physical content of Hooke's law.

Compressibility. A stretching (compression) of the bar length under the action of a stress X_x is associated to a compression (stretching) in the transverse direction, characterized by sizes w and h . Expression (3.14) indeed yields:

$$\frac{\Delta w}{w} = \frac{\Delta h}{h} = -\sigma \frac{\Delta \ell}{\ell}. \quad (3.17)$$

Consider now the case in which the material is acted on each face from the same pressure $p = F/A$. For the face normal to x -axis one has

$$\frac{\Delta \ell_1}{\ell} = -\frac{p}{G}, \quad (3.18)$$

whereas along the transverse directions y and z one has:

$$\frac{\Delta h_1}{h} = \frac{\Delta w_1}{w} = -\frac{p}{G}. \quad (3.19)$$

These compressions along the transverse directions correspond to the stretching along x -axis

$$\frac{\Delta \ell_2}{\ell} = \sigma \frac{p}{G}, \quad \frac{\Delta \ell_3}{\ell} = \sigma \frac{p}{G}. \quad (3.20)$$

The overall compression $\Delta \ell/\ell$ along the x -axis is found by summing (3.18) and (3.20), that is:

$$\frac{\Delta \ell}{\ell} = -\frac{p}{G} (1 - 2\sigma). \quad (3.21)$$

Symmetry considerations along the remaining directions yield:

$$\frac{\Delta w}{w} = \frac{\Delta h}{h} = -\frac{p}{G} (1 - 2\sigma). \quad (3.22)$$

Since the initial volume is $V = lwh$, one has $\Delta V = wh\Delta l + lh\Delta w + wl\Delta h$ and therefore:

$$\frac{\Delta V}{V} = \frac{\Delta l}{l} + \frac{\Delta w}{w} + \frac{\Delta h}{h} = -3\frac{p}{G} (1 - 2\sigma). \quad (3.23)$$

Thus:

$$p = -B \frac{\Delta V}{V}, \quad (3.24)$$

Concept

A pressure acted on a material results into a volume change ΔV . In the limit of small deformations and stresses, the overall relative volume change $\Delta V/V$ of a material subjected to pressure p has a linear-law dependence on pressure, according to a proportionality coefficient B^{-1} . This coefficient measures the material response to a compression.

Definition

The coefficient B is called compressibility. As long as the Poisson coefficient is smaller than 0.5, as it usually is, B is intrinsically a positive quantity.

By inspection, one sees that the compressibility B is by definition expressed as

$$B = \frac{G}{3(1 - 2\sigma)}. \quad (3.25)$$

Thus, B depends on both G and σ .

Concept

The relative stretching or compression along different unit vectors of a material are all related to each other, independently of the direction along which the stretching or compression has been originated.

Notice that whenever $\sigma \rightarrow (1/2)^-$, $B \rightarrow \infty$. Since $\Delta V/V$ has to be a finite quantity, an infinitely large compressibility B implies an infinitely large pressure p . Viceversa, if the applied pressure is finite, an infinite compressibility implies a vanishing relative volume change $\Delta V/V \rightarrow 0$.

Definition

A body with infinite compressibility can be considered as a rigid body.

Shape elasticity. The definition of shape elasticity is now given.

Definition

Conditions can be given under which the body changes its shape while its volume remains unchanged. This is called shape elasticity.

Such a situation can be represented by inserting $X_x = 0$ and $Z_x \neq 0$ into (3.13), with $Z_x \equiv g = F_t/A$ a shear force F_t acting on surface A . One then finds $e_1 = e_2 = e_3 = e_{yz} = e_{xy} = 0$ and

$$e_{zx} \simeq \alpha = \frac{Z_x}{c_{44}}. \quad (3.26)$$

Concept

A shear stress F_t/A acted on a material results into a shear deformation $e_{zx} \simeq \alpha$. In the limit of small stresses, the shear deformation has a linear-law dependence on shear stress, according to a proportionality coefficient $\mu^{-1} \equiv c_{44}^{-1}$. This coefficient measures the material response to a twist.

Definition

The coefficient μ is called shear or rigidity modulus.

The same phenomenon can be described by the following geometrical construction. Consider a section of the unit cube, that is a square, and surround it with a larger square whose sides be perpendicular to the diagonals of the original smaller square. Apply now a suited compression on two opposite faces and a stretching on the two remaining opposite faces. A quantitative analysis of this situation can be carried out in much the same way as done to derive the compressibility coefficients, and yields the relation

$$\mu = \frac{G}{2(1 + \sigma)}. \quad (3.27)$$

Expression (3.27) implies the condition $\sigma > -1$, so that typically the Poisson coefficient is bound as $-1 < \sigma < 1/2$.

A few useful concepts can be drawn from the analysis above.

Concept

Knowledge of the microscopic coefficients c_{ij} results into the knowledge of the macroscopic parameters describing the material response to various kinds of stresses,

therefore Young modulus, Poisson coefficient and shear modulus, and in general the elastic coefficients corresponding to whatever complex static conditions. Second,

Concept

A distinction exists between volume and shape elasticity, this being found also in the dynamical behavior of crystal ions.

The next section introduces the fundamental equations describing elasticity in the simpler case of cubic crystals and illustrates selected applications of the concepts discussed so far.

Quick Questions

Q1. Can the condition $\sigma > 1/2$ be physically meaningful?

Answer - It could be meaningful, under the following precise condition. From (3.23), $\sigma > 1/2$ implies a relative volume increase under the action of a compression. From (3.25), a negative compressibility would result, namely a compressed system would respond to a compression not contrasting it. This is physically meaningful, provided that an equilibrium is reached at some point which prevents the system from implosion.

Such a situation is experimentally realized for example in confined quantum atomic gases cooled down to nanokelvin temperatures below the threshold for Bose-Einstein condensation, when attractive interactions between atoms are set in: the condensate undergoes shrinking, until the finite size of the atoms hampers it any further.

Q2. Is a negative shear modulus μ meaningful?

Answer. No. It would imply that a surface subjected to a shear stress would slide in direction opposite to the shear stress.

Q3. Two materials have different Poisson coefficients: material A has a vanishing σ , material B has a finite $0 < \sigma < 1/2$. Both materials are subjected to the same stress per unit area F/A . Which one do you predict would undergo the larger deformation $\Delta\ell/\ell$?

Answer. According to Hooke's law, $\Delta\ell/\ell = G^{-1}F/A$, with $G = c_{11} - 2c_{12}\sigma$. Thus, material A has a larger rigidity G than material B. Thus, material A is expected to undergo a smaller relative deformation.

3.2.2 Dynamics of continuous media and elastic wave propagation

Consider an elementary infinitesimal cube with size $dx dy dz$ subjected to different forces acting on opposite faces. These forces can be calculated as follows. Consider the forces X_x and $-X_x$ acting along the x direction: the net force can be obtained after a first order Taylor expansion with respect to position around the cube center, that is:

$$\left(X_x + \frac{dx}{2} \frac{\partial X_x}{\partial x}\right) dy dz - \left(X_x - \frac{dx}{2} \frac{\partial X_x}{\partial x}\right) dy dz = \frac{\partial X_x}{\partial x} dx dy dz. \quad (3.28)$$

Let now apply the same procedure as that used to approximate the x -component of the net shear force acting on the surfaces perpendicular to axes y and z . One has

$$\frac{\partial X_y}{\partial y} dx dy dz, \quad \frac{\partial X_z}{\partial z} dx dy dz.$$

In this elementary cube with mass density ρ , the Newton equations for the displacements u , v , and w along x , y , and z axes respectively, can be expressed as

$$\begin{aligned} \rho \frac{\partial^2 u}{\partial t^2} &= \frac{\partial X_x}{\partial x} + \frac{\partial X_y}{\partial y} + \frac{\partial X_z}{\partial z}, \\ \rho \frac{\partial^2 v}{\partial t^2} &= \frac{\partial Y_x}{\partial x} + \frac{\partial Y_y}{\partial y} + \frac{\partial Y_z}{\partial z}, \\ \rho \frac{\partial^2 w}{\partial t^2} &= \frac{\partial Z_x}{\partial x} + \frac{\partial Z_y}{\partial y} + \frac{\partial Z_z}{\partial z}. \end{aligned} \quad (3.29)$$

The significant meaning of these equations of motion can be evidently extracted by analyzing the simple case of a cubic-symmetric system. In this case, one uses the symmetries relating the coefficients c_{ij} . Considering the constitutive relations (3.12), one then finds

$$\begin{aligned}\rho \partial_{tt} u &= c_{11} \partial_{xx} u + c_{44} (\partial_{yy} u + \partial_{zz} u) + (c_{12} + c_{44}) (\partial_{xy} v + \partial_{xz} w) \\ \rho \partial_{tt} v &= c_{11} \partial_{yy} v + c_{44} (\partial_{xx} v + \partial_{zz} v) + (c_{12} + c_{44}) (\partial_{xy} u + \partial_{yz} w) \\ \rho \partial_{tt} w &= c_{11} \partial_{zz} w + c_{44} (\partial_{yy} w + \partial_{xx} w) + (c_{12} + c_{44}) (\partial_{yz} v + \partial_{xz} u),\end{aligned}\quad (3.30)$$

where $\partial_{tt} \equiv \partial^2 / \partial t^2$.

The shear coefficient $c_{44} = \mu$ has already been analyzed. Furthermore, if $c_{11} - c_{12} = 2c_{44}$ and setting $c_{12} = \lambda$, one then finds $c_{11} = \lambda + 2\mu$, so that (3.30) assume a quite significant form. For example, the first among (3.30) yields

$$\begin{aligned}\rho \partial_{tt} u &= (\lambda + 2\mu) \partial_{xx} u + \mu (\partial_{yy} u + \partial_{zz} u) + (\lambda + \mu) (\partial_{xy} v + \partial_{xz} w) \\ &= (\lambda + \mu) \partial_x (\nabla \cdot \mathbf{s}) + \mu \nabla^2 u.\end{aligned}\quad (3.31)$$

Along these lines, one can conclude that the condition $c_{11} - c_{12} = 2c_{44}$ corresponds to an isotropic material. The remaining two equations are manipulated in similar manner. It turns out useful to define the displacement vector $\mathbf{s} = u\hat{\mathbf{x}} + v\hat{\mathbf{y}} + w\hat{\mathbf{z}}$ out of its u , v , and w components. With this definition, all the three (3.30) can thus be expressed in the compact and meaningful form

$$\rho \partial_{tt} \mathbf{s} = (\lambda + \mu) \nabla (\nabla \cdot \mathbf{s}) + \mu \nabla^2 \mathbf{s}, \quad (3.32)$$

The condition $c_{11} - c_{12} = 2c_{44}$ has transformed the equations of motion (3.30) into (3.32) describing the behavior of an isotropic medium after deformation. The solutions of (3.32) can be classified as follows:

1. Irrotational case, namely with $\nabla \wedge \mathbf{s} = 0$. Set $\mathbf{s} = \nabla \phi$ so that $\nabla \wedge \mathbf{s} = 0$. One then finds

$$\rho \partial_{tt} \phi = (\lambda + 2\mu) \nabla^2 \phi; \quad (3.33)$$

2. Rotational case, namely with $\nabla \cdot \mathbf{s} = 0$. Set $\mathbf{s} = \nabla \wedge \boldsymbol{\psi}$ instead, so that $\nabla \cdot \mathbf{s} = 0$. One then finds

$$\rho \partial_{tt} \boldsymbol{\psi} = \mu \nabla^2 \boldsymbol{\psi}. \quad (3.34)$$

Expressions (3.33)-(3.34) represent wave equations. The following general physical concept expressed by (3.30) thus emerges from the specific example of cubic symmetry:

Concept

The dynamical behavior of elementary portions of a continuous medium corresponds to elastic (mechanical) wave propagation.

Selected solutions of (3.30) are now being discussed. The analysis remains in fact within cubic-symmetric systems, where equations are simpler and easier to be handled and read.

Specific cases Seek solutions of the type $\mathbf{s}(\mathbf{r}, t) = \mathbf{s}_0 e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}$ with initial conditions $\mathbf{s}_0 \equiv (u_0, v_0, w_0)$. Equation (3.30) becomes:

$$\begin{aligned}\rho \omega^2 u_0 &= [c_{11} k_x^2 + c_{44} (k_y^2 + k_z^2)] u_0 + (c_{12} + c_{44}) (k_x k_y v_0 + k_x k_z w_0), \\ \rho \omega^2 v_0 &= [c_{11} k_y^2 + c_{44} (k_x^2 + k_z^2)] v_0 + (c_{12} + c_{44}) (k_x k_y u_0 + k_y k_z w_0), \\ \rho \omega^2 w_0 &= [c_{11} k_z^2 + c_{44} (k_x^2 + k_y^2)] w_0 + (c_{12} + c_{44}) (k_x k_z u_0 + k_y k_z v_0). \quad (3.35)\end{aligned}$$

These equations are a homogeneous linear set of equations in the unknown quantities u_0, v_0, w_0 . Setting to zero the determinant of the coefficients, three relations between ω and k_x, k_y, k_z are found.

Definition

The relation between the wave frequencies ω and the components k_x, k_y and k_z of \mathbf{k} can be found. These relations are called dispersion relations.

Within the degree of approximation embodied in the Taylor expansions performed in the present derivation, these relations are linear.

No definite angular relations exist in general between vectors \mathbf{k} and \mathbf{s}_0 . When specific relations set in instead, special situations occur. For example:

- Case with $k_x \neq 0$ and $k_y, k_z = 0$. The solutions can be derived, and are:
 1. One solution with $\omega/k_x = \sqrt{c_{11}/\rho}$, $v_0 = w_0 = 0$ and $u_0 \neq 0$. This represents a longitudinal wave, with c_{11} the compressibility coefficient determining the wave group velocity.
 2. Two solutions with $\omega/k_x = \sqrt{c_{44}/\rho}$, one with $u_0 = w_0 = 0$ and $v_0 \neq 0$, and one with $u_0 = v_0 = 0$ and $w_0 \neq 0$. These represent two transverse waves, with c_{44} the shape-elasticity coefficient.
- Case with $k_z = 0$ and $k_x = k_y = k/\sqrt{2}$. The solutions are:
 1. A first transverse wave with $\omega/k = \sqrt{c_{44}/\rho}$, $u_0 = v_0 = 0$ and $w_0 \neq 0$.
 2. A second transverse wave with $\omega/k = \sqrt{(c_{11} - c_{12})/2\rho}$, $w_0 = 0$ and $u_0 = -v_0$.
 3. One longitudinal wave with $\omega/k = \sqrt{(c_{11} + c_{12} + 2c_{44})/2\rho}$, $w_0 = 0$ and $u_0 = v_0$.
- Case with $k_x = k_y = k_z = k/\sqrt{3}$. The solutions are:
 1. Two transverse waves with $\omega/k = \sqrt{(c_{11} - c_{12} + c_{44})/3\rho}$, one with $w_0 = 0$ and $u_0 = -v_0$, and one with $v_0 = 0$ and $u_0 = -w_0$.

2. One longitudinal wave with $\omega/k = \sqrt{(c_{11} + 2c_{12} + 4c_{44})/3\rho}$, and $u_0 = v_0 = w_0$.

Problem 3.3 shows how the condition $c_{11} - c_{12} = 2c_{44}$ implies that all the transverse waves share the same velocity $\sqrt{c_{44}/\rho}$, whereas all the longitudinal waves share the same velocity $\sqrt{c_{11}/\rho}$, different than the transverse wave velocity. This result is also found by inspection of the wave equations (3.33) and (3.34): the corresponding solutions indeed possess dispersion relations $\omega/k = \sqrt{(\lambda + 2\mu)/\rho}$ and $\omega/k = \sqrt{\mu/\rho}$. These precisely are the dispersion relations for longitudinal and transverse waves, respectively.

Table 3.1 lists the elastic constants of selected materials with cubic structure. Already at the present depth of discussion, it is intuitive that the response of the system to stresses in general depends on temperature. The values of the listed elastic constants refer to room temperature. For completeness, the corresponding material densities are also reported.

Table 3.1 Elastic constants c_{11} , c_{12} and c_{44} at room temperature of the selected materials with cubic structure, listed in first column. The corresponding densities are reported in the last column [14]

| Crystal | Elastic coefficients ($\times 10^{12}$ dine/cm ²) | | | Density (g/cm ³) |
|-------------|---|----------|----------|---------------------------------|
| | c_{11} | c_{12} | c_{44} | |
| Li | 0.134 | 0.113 | 0.096 | 0.53 |
| Na | 0.076 | 0.063 | 0.043 | 0.97 |
| K | 0.057 | 0.032 | 0.019 | 0.86 |
| Al | 1.08 | 0.62 | 0.283 | 2.70 |
| Cu | 1.69 | 1.22 | 0.753 | 8.96 |
| Ag | 1.23 | 0.92 | 0.453 | 10.49 |
| Au | 1.90 | 1.61 | 0.423 | 19.30 |
| C (diamond) | 10.40 | 1.70 | 5.50 | 3.52 |
| Si | 1.65 | 0.64 | 0.792 | 2.33 |
| Ge | 1.29 | 0.48 | 0.671 | 5.33 |
| LiF | 1.2 | 0.46 | 0.635 | 2.64 |
| NaCl | 0.491 | 0.128 | 0.128 | 2.17 |
| KCl | 0.405 | 0.069 | 0.063 | 1.99 |
| RbCl | 0.364 | 0.063 | 0.047 | 2.80 |
| CsBr | 0.307 | 0.084 | 0.075 | 4.45 |

Quick Questions

Q4. Consider a cubic-symmetric material with either negative c_{11} or negative c_{44} . What does it imply for the longitudinal and transverse wave propagation?

Answer. The propagation velocity of either longitudinal or transverse waves, respectively, would be imaginary. As a result, if the frequency ω were real, the wavenumber would be imaginary. As a consequence, the displacement amplitudes would be exponentially reduced along selected directions, and amplified along the opposite ones. If instead the wavenumber k were real, the frequencies would result to be imaginary, so that the wave amplitude would be exponentially amplified or reduced while time flows.

Q5. Equation (3.32) possesses either longitudinal or transverse solutions.

Is this true also for other directions in cubic-symmetric materials?

Answer. No, it is not a general result. Along directions different from those considered in the examples, solutions are not in general either longitudinal or transverse.

Application. Bending of a rod with one free end. The theoretical results derived so far in the present section can be usefully applied to real situations. In particular, the model of a rod with one fixed and one free end subjected to bending is useful to treat and understand the dynamics of nanowires with one bound end. The reasoning closely follows the pedagogical presentation by Feynman [3]. Fig. 3.1(a) sketches the model, where the bent rod is assumed to be shaped according to an arch of circumference with radius R . In order to simplify the calculations, the assumption is put forward that the linear size d of the rod section be much smaller than R ($d \ll R$). Fig. 3.1(b-c) sketches instead a small portion of the bent rod, along with traction and compression forces that the remaining parts of the rod exert on the infinitesimal element under study. Traction and compression forces on an elementary rod portion have opposite signs. Thus, a surface is to exist that results to be neutral with respect to deformations, namely not deformed. The amplitude of the deformation originated by traction is as much larger as longer is the distance y of the elementary portion from the neutral surface. The deformation amplitude is also as much smaller as longer is the radius R of the bent rod. In the linear limit of small bending amplitudes, these considerations are represented in mathematical form as:

$$\frac{\Delta l}{l} = \frac{y}{R}.$$

The relative stretching $\Delta l/l$ enters the law (3.16), that rules the amount of stretching due to traction forces F per unit surface acting on a surface portion ΔA . Using (3.16), one then finds

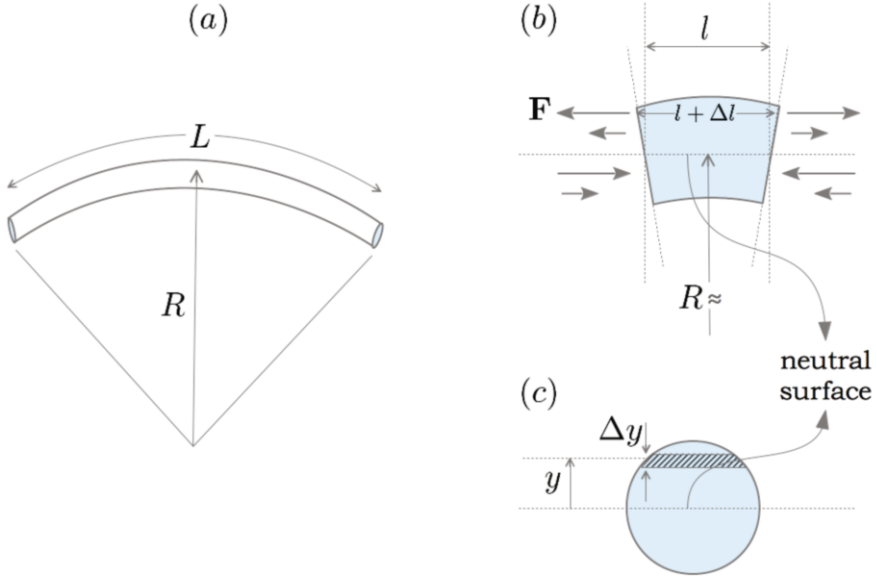


Fig. 3.1 Elastic behavior and wave propagation in continuous media [3]. (a) Model for a rod bent in the shape of a circumference with radius R . In the model, the linear transverse size $d \ll R$ is negligible and small bending amplitudes are considered. (b-c) Zoom on an elementary portion of the rod with length Δl . The traction and compression forces exerted on it by the neighboring portions of the rod are sketched. The neutral surface is shown, as discussed in the main text

$$\frac{\Delta F}{\Delta A} = G \frac{\Delta l}{l} = G \frac{y}{R}. \quad (3.36)$$

Along these lines, the relevant concept in this bending game is the net torque acting on the elementary portion. In order to calculate it, a pole has to be chosen, that is conveniently located at the neutral surface. The size of the torque M of force F with respect to the neutral surface is

$$M = \int y dF = \frac{G}{R} \int y^2 dA = \frac{G}{R} I, \quad (3.37)$$

with $I \equiv \int y^2 dA$ a geometrical factor. As a result, the bending torque strength increases with increasing I . In essence, in construction contexts a beam is to be conveniently shaped to have large I values, for best control of bending.

The result (3.37) can be specified in the case of a rod with one end fixed and one end free, on which a force W acts to as schematically represented in Fig. 3.2. Assume for simplicity that the rod free end has negligible mass. The presence of the force W introduces a complexity, since the curvature radius corresponding to the given rod point varies along the rod. Techniques of differential geometry help to link the curvature radius to the rod bending $z(x)$ at point x . One has:

$$\frac{1}{R} = \frac{\frac{d^2 z(x)}{dx^2}}{\sqrt{1 + \left[\frac{dz(x)}{dx}\right]^2}} \simeq \frac{d^2 z(x)}{dx^2},$$

where the second relation holds in the limit $|dz/dx| \ll 1$. Combining this expression with $M = W(L - x)$, expression (3.37) reads:

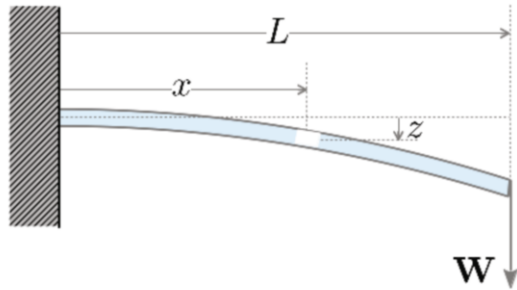
$$\frac{d^2 z(x)}{dx^2} = \frac{W}{GI} (L - x). \quad (3.38)$$

When integrating (3.38), the boundary conditions are to be given. Under the present circumstances, one has $z(0) = 0$ and $dz(x)/dx|_{x=0} = 0$, so that integration yields

$$W = \frac{3GI}{L^3} z(L), \quad (3.39)$$

with L the overall rod length. In essence, bending $z(L)$ at the rod end and applied force are proportional to each other, the proportionality coefficient being determined by the Young modulus and the geometrical factor I .

Fig. 3.2 Elastic behavior and wave propagation in continuous media [3]. Model to study the bending of a rod with one end fixed and one end free, on which a force \mathbf{W} is acted on. The model approximations are as in Fig. 3.1(a). In addition, the mass of the rod is assumed to be negligible



These results can be applied to the case example of carbon nanowires fixed at one end. These can be modeled as continuous materials composed of a cylindrically shaped ring with external radius a and internal radius b . Under these geometrical conditions, the neutral surface is a plane containing the principal symmetry axis of the cylinder, that is normal to the bending direction. The geometrical factor results to be:

$$I = \int_b^a dr \int_0^{2\pi} d\theta r^3 \sin^2 \theta = \frac{\pi}{4} (a^4 - b^4).$$

As a consequence, the proportionality coefficient between W and $z(L)$ in (3.39) is

$$c_0 = \frac{3}{4} \frac{\pi G (a^4 - b^4)}{L^3}. \quad (3.40)$$

The complete theory for the bending modes of a rod provides also the energy U_n of n -th mode, that is:

$$U_n = \frac{1}{2} c_n u_n^2, \quad (3.41)$$

$$c_n = \frac{\pi \beta_n^4 G (a^4 - b^4)}{16L^3}, \quad (3.42)$$

with u_n the oscillation amplitude and the coefficients β_n determined after solution of the implicit equation $\cos(\beta_n) \cosh(\beta_n) + 1 = 0$. For example, one obtains $\beta_0 \simeq 1.8751$, $\beta_1 \simeq 4.6941$, \dots . The oscillation frequency is:

$$\omega_n = \frac{\beta_n^2}{2L^2} \sqrt{\frac{G(a^2 + b^2)}{\rho}}, \quad (3.43)$$

where the linear density $\rho = M/L$ has been introduced. It is seen that the structure for c_n in the complete theory represented in (3.42) is the same as that for c_0 in the approximated analysis (3.40), except for the different numerical factors in front. For the fundamental mode, the complete theory in (3.42) yields $\beta_0^4/16 = 0.773$, not far from the result of the approximated factor $3/4 = 0.75$ in (3.40).

The model can be used to infer the Young modulus of the nanowire. Two alternative methods are discussed in the following. In a first method, deflections of the free nanowire end are measured by optical means, while temperature is varied. Under common conditions $k_b T \gg \hbar \omega_n$, one has $\langle U_n \rangle = k_b T/2$. Thus, a measure of the average oscillation amplitudes as functions of temperature T should do the job. The modes are indeed expected to be distributed according to Boltzmann statistics with average squared amplitude linearly dependent on T . If this is the case, a fit of experimental data aimed to extract the slope of the linear behavior, provides the value for G . Fig. 3.3 displays experimental data for bending oscillations of thin carbon nanowires fixed at one end [4]. The average squared oscillation amplitudes $\langle u^2 \rangle$ are represented while temperature is varied. The linear behavior is clearly visible. Data analysis also allows to conclude that nanowires of the same tickness have larger Young modulus when shorter, G being indeed a decreasing function of the length.

An alternative method is to measure the free-end deflection originated by the action of an electrical field, as a function of the electrical potential difference. In this type of experiment, carbon nanowires are placed between two electrodes at potential difference V_s . A charge is induced on the nanowire, that accumulates at the free end and amounts to a value proportional to V_s . The acting force in turn, is proportional to the potential difference. Thus, the size of the wire deflection is expected to be proportional to the squared potential difference. Fig. 3.4 displays experimental data for the free-end static deflection of nanowires placed between two electrodes, as a function of the applied potential difference at the electrodes V_s [5]. The quadratic behavior is clearly visible. In the same experiment, resonant oscillations have been obtained after finely tuning the frequency of a time-dependent applied $V_s(t)$ with respect to the frequencies of normal modes. The same conclusions on the dependence

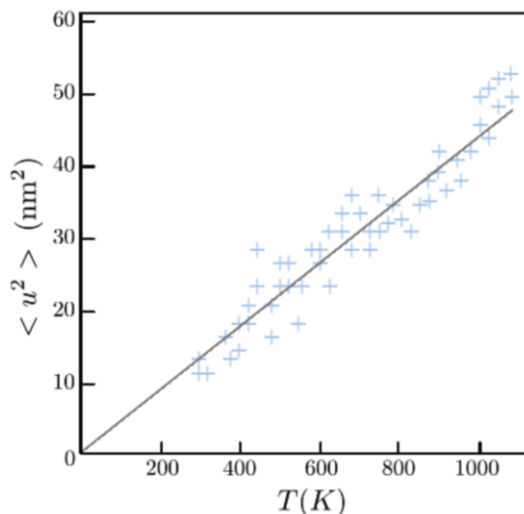


Fig. 3.3 Elastic behavior and wave propagation in continuous media. Bending of free-standing carbon nanowires after application of a force as sketched in Fig. 3.2. Experimental data are shown, representing the average squared oscillation amplitude $\langle u^2 \rangle$ as a function of temperature T . The predicted linear behavior is fitted and is clearly visible as represented by the solid line. The resulting G value is $G = 0.044 \pm 0.0018 \text{ nm}^2\text{K}^{-1}$. The data are the result of a number of measurements for statistical reasons [4]. Details on the measurement technique are referred to the original work by Treacy et al. [4]

of G on the nanowire thickness are obtained, as seen by comparing data referring to different nanowires and represented by filled and open circles in the figure. Problem 3.4 elaborates on a model that explains how the Young modulus of nanometric systems occurs to be larger than that of corresponding macroscopic systems.

Quick Questions

Q6. Consider a stele, in the form of a macroscopic piece of steel with cylindrical shape. The stele is vertically placed and grounded. Describe its dynamical behavior under the action of strong winds with randomly variable directions.

Answer. The analysis is precisely the same as that drawn for a carbon nanowire. Thus...

3.3 Piezoelectricity and piezomagnetism

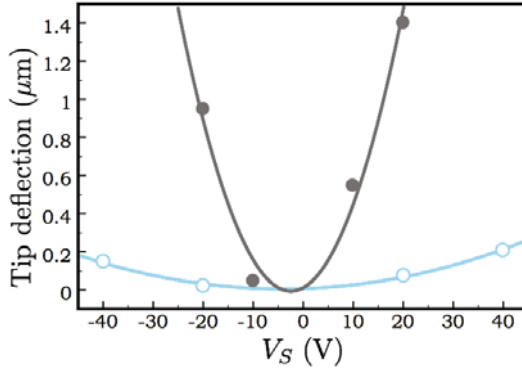


Fig. 3.4 Elastic behavior and wave propagation in continuous media. Bending of free-standing carbon nanowires placed between two electrodes after applying a potential difference V_S . Experimental data are shown, representing with symbols the static deflection of the free end as a function of potential difference V_S . Different symbols refer to nanowires with different thickness and length. Filled circles refer to thinner and longer nanowires, with respect to the case represented by open circles. The predicted quadratic behavior is fitted and is clearly visible as represented by the solid line. The same dependence of G on the nanowire thickness is found, as that extracted from Fig. 3.3. Details on the measurement technique are referred to the original work by Poncharal et al. [5]

3.3.1 General facts

The elasticity phenomena discussed in the previous section originate from mechanical stresses, and have been found to be characterized by symmetric coefficients in the linear relations between the six stress and six strain quantities. Materials exist, in which elastic response can be excited by non-mechanical stresses and in which mechanical stresses can produce a non-mechanical response. In piezoelectric materials, stresses originate macroscopic electric polarization and in piezomagnetic materials they originate a macroscopic magnetization.

Not all materials can be piezoelectric: one special condition is at least necessary. Certainly, those materials having the parity inversion operator in the space crystalline group, cannot be piezoelectric: if they were so, the electric polarization \mathbf{P} should change sign under inversion, leading to $\mathbf{P} = 0$. In essence, piezoelectricity is strictly related to crystal symmetry and to the symmetry of the base atoms. On the other hand, piezoelectricity and polar properties are not necessarily related phenomena. Piezoelectric materials often occur to be ferroelectric as well, that is to spontaneously show up a macroscopic electric polarization in analogy to ferromagnets. The most common piezoelectric material is quartz, SiO_2 .

To see how piezoelectricity works at least at a macroscopic level, it is useful to set the corresponding constitutive equations. In a manner similar to (3.11), the equations now relate the components P_i of the electric polarization to those of the electric field E_i via the deformation parameters e_{kl} :

$$P_i = \sum_k \chi_{ik} E_k + \sum_{kl} \gamma_{ikl} e_{kl}, \quad (3.44)$$

where χ_{ik} are the components of the electric susceptibility tensor and γ_{ikl} are the piezoelectric coefficients. The meaning of γ_{ikl} is easily caught when setting $\mathbf{E} = 0$ and all but one e_{kl} to vanishing values: γ_{ikl} represents the polarization P_i at zero electric field. On the other hand, in the case of constant displacement vector \mathbf{D} , the relations

$$\delta e_{kl} = \bar{\gamma}_{kli} \delta E_i \quad (3.45)$$

state that a variation of the electric field δE_i determines a change of the deformation parameter δe_{kl} . Equation (3.45) represents the inverse piezoelectric effect. Knowledge of the nonzero piezoelectric coefficients is essential to engineer piezoelectric devices, and can in general be built up by analyzing the crystal symmetries.

The description of piezomagnetism closely follows that of piezoelectricity: after substituting magnetization \mathbf{M} and magnetic field \mathbf{B} in place of polarization \mathbf{P} and electric field \mathbf{E} it is possible to repeat the above discussion and describe in such way the phenomenon of the piezomagnetism. The treatment of these phenomena is kept within the present macroscopic framework in this textbook, since a microscopic treatment would require a too high degree of complexity.

3.3.2 Applications

Piezoelectric materials are widely used to manufacture transducers, that are devices transforming electrical into mechanical energy (and signals) and back. One among the first engineered devices is the sonar, to spot far objects under the sea. In the sonar, an electric signal makes the piezoelectric material to vibrate, so that these vibrations may originate mechanical pressure waves in surrounding water. The water waves propagate until they reach the object, from which they are reflected back. The reflected waves reach the sonar back, putting it in mechanical vibration, that is in turn transformed in a new electric signal to be analyzed. In essence, the sonar works first as a speaker and then as a microphone, which are indeed other quite common applications of piezoelectric properties [1].

One more application is in precise clocks manufacturing. Quartz, that is SiO_2 crystals effectively serve the purpose. As already discussed in Chap. 1, quartz is hard (and brittle), so that is characterized by especially stiff elastic restoring forces. Therefore, when struck, quartz vibrates at large frequencies, that helps in reaching higher precisions in frequency and thus time measurement. In addition, quartz is characterized by very small dissipative forces and by a quite negligible volume changes in response to gradient temperatures. But above all, quartz is piezoelectric, so that the vibrations can be started and read back by means of simple (and tiny) electrical circuitry. For all these reasons, modern wrist clocks contain tiny quartz crystals, as they are as portable as mechanical wrist clocks, but sharper. Notice that in order to preserve battery life, the fundamental vibration frequency of the quartz crystal is to be as low as possible: thus, quartz crystals for clocks are shaped as

tiny tuning forks, where the mass is indeed reduced by excavating some mass away. Common watch crystals vibrate at 2^{15} Hz. Lower frequencies and longer battery lives could in principle be reached, but are not in the practice: vibration frequencies would indeed become audible, below the ≈ 20000 Hz upper threshold of human hearing system [1].

Considering the usefulness of piezoelectric-based devices in everyday life, it is no surprise that current applied research is quite interested in engineering materials with large piezoelectric coefficients.

3.4 Microscopic picture: normal modes

Sec. 3.2 has investigated the static and dynamical behaviors of materials subjected to stresses, within a framework where their structure is considered as a continuum. The present section faces the same problem, but reverting back to the discrete structure analyzed in Chaps. 1 and 2. There, atoms and molecules of the crystal structure are point-like elements, whose dynamics is governed by Schrödinger equations within the Born-Oppenheimer approximation. In this approximation, the ion vibrational state energies are much smaller as compared to typical electron energies, and the eigenvalues of the electronic structure problem enter as interaction potentials between atoms or molecules. Within a harmonic approximation, these considerations lead to a tractable solution of the full problem. The above framework, so far classical, is then amenable to be adapted and extended to quantum treatment. The present section is aimed to develop the quantum treatment of ion oscillations around their equilibrium lattice positions, within the Born-Oppenheimer approximation.

The goal is pursued by the following procedure. First, the eigenvalues $\varepsilon_n(\mathbf{R})$, considered as functions of \mathbf{R} , are expanded in series around the position \mathbf{R}^0 corresponding to equilibrium, and thus to a minimum of $\varepsilon_n(\mathbf{R})$. One then writes:

$$\varepsilon_n(\mathbf{R}) \simeq \Phi_0 + \frac{1}{2} \sum_{J,K} \phi_{J,K} s_J s_K \equiv \Phi(\mathbf{R}), \quad (3.46)$$

$$\phi_{J,K} = \frac{\partial^2 \varepsilon_n(\mathbf{R})}{\partial R_{\alpha,i} \partial R_{\beta,j}} \Big|_{\mathbf{R}=\mathbf{R}^0} = \phi_{K,J}, \quad (3.47)$$

with s_J e s_K representing the ion displacements. Indexes J and K label both position and displacement components for each atom. Equation (3.47) corresponds to a harmonic approximation to the potential. It is a positive definite quadratic form in variables s_J . Notice that the constant term Φ_0 in (3.47) is precisely the quantity yielding the system cohesive energy discussed in Chap. 1.

Second, one solves the classical problem and eventually shows that the found solutions correspond as well to quantum solutions. The advantage of this procedure is to keep up the pace while building the connection between the macroscopic picture of Sec. 3.2 and the present microscopic understanding.

The classical equations of motion for the ions in the harmonic potential Φ in (3.47) are:

$$M_J \ddot{s}_J = -\nabla_J \Phi = -\sum_K \phi_{J,K} s_K, \quad J = 1, \dots, 3N \quad (3.48)$$

with M_J the mass of ion J and N the total number of ions. The short notation \ddot{s}_J stands here for second derivative with respect to time. Introducing the convenient variables $q_J = \left(\sqrt{M_J/M}\right) s_J$, the set of equations (3.48) becomes:

$$\ddot{q}_j = -\sum_k D_{j,k} q_k, \quad (3.49)$$

with $M = \sum_J M_J / (3N)$ the average mass of the atoms in the system and $D_{J,K} \equiv \phi_{J,K} / \sqrt{M_J M_K} = D_{K,J}$. Thus, a set of $3N$ second order differential equations with constant coefficients results. This is solved by seeking for $6N$ independent particular solutions of the form $q_J = Q_J e^{-i\omega t}$. Substitution of this ansatz into (3.49), the following set of $3N$ homogeneous linear algebraic solutions is obtained:

$$\omega^2 Q_J = \sum_K D_{J,K} Q_K. \quad (3.50)$$

The existence of non trivial solutions is conditioned by having a vanishing secular determinant with rank $3N$, that is by the condition

$$|\omega^2 \delta_{J,K} - D_{J,K}| = 0. \quad (3.51)$$

By inspection, the symmetry relation $D_{J,K} = D_{K,J}$ holds. The determinant is associated to a symmetric quadratic form that is positive definite, thus the corresponding determinant equation has $3N$ real positive solutions. The frequencies ω_α^2 are then obtained, with $\alpha = 1 \dots 3N$. The $3N$ particular solutions of (3.50) can thus be expressed in the form

$$q_{J,\alpha}(t) = Q_{J,\alpha} e^{-i\omega_\alpha t},$$

with $J = 1 \dots 3N$ and α given. The general solution to the set of (3.50) is obtained by forming the linear combination

$$q_J(t) = \sum_\alpha (A_\alpha e^{-i\omega_\alpha t} + B_\alpha e^{i\omega_\alpha t}) Q_{J,\alpha}, \quad (3.52)$$

in terms of the $6N$ arbitrary integration constants A_α and B_α , to be determined after the knowledge of the initial conditions. The eigenvectors calculated from (3.50) satisfy the orthogonality and completeness conditions

$$\begin{aligned} \sum_J Q_{J,\alpha} Q_{J,\alpha'} &= \delta_{\alpha,\alpha'}, & \text{orthogonality,} \\ \sum_\alpha Q_{J,\alpha} Q_{K,\alpha} &= \delta_{J,K}, & \text{completeness.} \end{aligned} \quad (3.53)$$

The variables q_J and a_α are related by a one-to-one mapping. To see this, one can first write (3.52) in the form

$$q_J(t) = \sum_\alpha a_\alpha(t) Q_{J,\alpha}, \quad (3.54)$$

and then invert (3.54). To this aim, one multiplies each side of the equation by $Q_{J,\alpha}$ and sums over J . One then obtains

$$\sum_J Q_{J,\alpha} q_J(t) = \sum_J Q_{J,\alpha} \sum_{\alpha'} a_{\alpha'}(t) Q_{J,\alpha'} = \sum_{\alpha'} a_{\alpha'}(t) \delta_{\alpha,\alpha'} = a_{\alpha}(t), \quad (3.55)$$

where the last two equalities are a consequence of (3.53).

Definition

The $3N$ independent variables $a_{\alpha}(t)$ are called normal variables. They result from linear combinations of the displacements and thus are by construction collective variables, which can be conveniently used to describe the system dynamics in place of the displacement variables.

Appendix 3.13 illustrates the convenience of this choice, since both kinetic K and potential Φ energies can be written as quadratic normal forms of these new collective variables. In summary, one has:

$$\begin{aligned} K &= \frac{M}{2} \sum_{\alpha} \dot{a}_{\alpha}^2, \\ \Phi &= \frac{M}{2} \sum_{\alpha} \omega_{\alpha}^2 a_{\alpha}^2. \end{aligned} \quad (3.56)$$

As a result, the system Hamiltonian takes an especially appealing form in the new collective variables:

$$H = \frac{M}{2} \sum_{\alpha} (\dot{a}_{\alpha}^2 + \omega_{\alpha}^2 a_{\alpha}^2) \quad (3.57)$$

In essence,

Concept

The collective dynamics of a system of N ions moving around their equilibrium lattice positions can be described by the Hamiltonian for a set of independent harmonic oscillators.

This is true within the second-order harmonic approximation to the effective potential $\varepsilon_n(\mathbf{R})$, which in turn results from the Born-Oppenheimer approximation.

The procedure to obtain H in (3.57) is useful also to easily head towards the quantum treatment. Indeed, the transformation (3.54) is unitary, and in quantum mechanics a unitary operation leaves dynamical variables unchanged. In classical mechanics the motion described by variables $a_{\alpha}(t)$ are the so-called normal modes with eigenfrequencies given by ω_{α}^2 . In quantum mechanics, normal variables and their time derivatives play the role of conjugates variables of the harmonic oscillator.

Concept

The quantization of the harmonic oscillator normal variables defines the concept of elementary vibrational excitation, that is a phonon.

As already seen, the computation of the frequencies ω_α emerges from the zeros of a polynomial function of order $3N$. For a simple molecule N can be as small as $\simeq 10$, but for a nanostructure can be as large as $\simeq 10^3$ or as huge as $\sim 10^{23}$ for a solid. The same treatment is in fact valid for whatever aggregate of atoms, be it a molecule, nanostructure or solid, and the problem complexity is transferred into the determination of the coefficients $\phi_{J,K}$ and of the normal modes frequencies ω_α .

Quick Questions

Q7. Write the Hamiltonian for a biatomic heteronuclear molecule, constrained to move along a straight line, whose atoms interact through an elastic force with stiffness κ . Which are the frequencies of the normal modes?

Answer. The molecule has two atoms and has to have two normal modes. The molecule center of mass is constrained to move on a straight line. No net forces act on it, so that the center of mass moves with constant velocity. This motion can be viewed as a vibrational mode with zero frequency. The second mode is obtained after writing the equations of motion

$$\begin{aligned} M_1 \ddot{x}_1 &= \kappa(x_2 - x_1), \\ M_2 \ddot{x}_2 &= -\kappa(x_2 - x_1). \end{aligned} \quad (3.58)$$

and executing respectively side by side the sum and the difference of the two equations. The sum yields $(M_1 \ddot{x}_1 + M_2 \ddot{x}_2) = M \ddot{x}_{CM} = 0$, which indeed describes the first mode with M the total mass and x_{CM} the coordinate of the center-of-mass vibrating at zero frequency. The difference yields $\ddot{x}_2 - \ddot{x}_1 = -\kappa(1/M_2 + 1/M_1)(x_2 - x_1)$ or $\ddot{x}_{rel} = -(\kappa/\mu)x_{rel}$, which describes the second mode with frequency $\omega = \sqrt{\kappa/\mu}$, μ being the reduced mass and x_{rel} the relative coordinate.

3.5 Vibrational modes of systems with finite number of atoms

Single-particle electronic excitations in the crystal have been studied in Chap. 2, starting from simple one-dimensional systems with finite size. The electronic excitations of infinite systems have been obtained by extending the reasoning with approximate boundary conditions. Finally, additional complexities have been introduced, such as those due to higher dimensionality and interactions. In similar

manner, single-particle vibrational excitations of the crystal are studied in the following, starting from simple one-dimensional (1D) systems with finite size, then evolving the results towards an infinite system but with different kinds of boundary conditions, and eventually introducing higher dimensions and anharmonic effects.

3.5.1 Linear chain with N identical atoms interacting with first neighbors via elastic forces

Selected examples of linear chains are here studied, with interactions that are elastic in nature, and limited to first neighbors. Different boundary conditions are considered, for the reasons that become evident in the course of the treatment.

3.5.2 Linear atomic chain with free edge atoms

In a way similar to the study of electronic structures, one begins with considering a finite 1D chain structured with N identical atoms with mass M placed at equilibrium positions and separated by a constant pace a . Each atom interacts only with its first neighbors, through elastic forces with stiffness κ . This assumption of short-range interactions is justified a posteriori in Chap. 4: in essence, electrons are able to efficiently screen the Coulomb interactions between the heavier, thus slower, ions. No other external nor internal forces are assumed to act on the system. The first and last atoms, placed at the two opposite ends of the chain, are free. The equations of motion for the displacements s_n of n -th atom are:

$$\begin{aligned} M\ddot{s}_1 &= \kappa(s_2 - s_1), & \text{for } n = 1, \\ M\ddot{s}_n &= \kappa(s_{n+1} - s_n) + \kappa(s_{n-1} - s_n), & \text{for } n = 2, \dots, N-1, \\ M\ddot{s}_N &= \kappa(s_{N-1} - s_N), & \text{for } n = N. \end{aligned} \quad (3.59)$$

Different boundary conditions for the edge atoms, such as they being held at rest, require the introduction of additional suited forces, as discussed later on in the section.

The solutions to the set of equations (3.59) can be sought for in the form $s_n(t) = Q_n e^{-i\omega t}$, so that the linear set of equations (3.59) gets the form:

$$\begin{aligned} -M\omega^2 Q_1 &= \kappa(Q_2 - Q_1), & \text{for } n = 1, \\ -M\omega^2 Q_n &= \kappa(Q_{n+1} - Q_n) + \kappa(Q_{n-1} - Q_n), & \text{for } n = 2, \dots, N-1, \\ -M\omega^2 Q_N &= \kappa(Q_{N-1} - Q_N) & \text{for } n = N. \end{aligned} \quad (3.60)$$

Problem 3.5 expands the detailed calculations to solve (3.60). In summary, one finds that the equations on the second line are solved by $Q_n = A \sin(n\varphi) + B \cos(n\varphi)$, provided that

$$\omega^2 = \frac{2\kappa}{M} (1 - \cos \varphi). \quad (3.61)$$

for arbitrary values A , B and φ . The arbitrary constants A and B are determined from the boundary conditions, that are represented by the first and the last lines in (3.60). The phase φ is determined by requiring non trivial solutions for A and B . In particular, use of the result (3.61) yields

$$\begin{aligned} -A \sin \varphi + B(1 - \cos \varphi) &= 0, \\ \alpha A + \beta B &= 0, \end{aligned} \quad (3.62)$$

with

$$\begin{aligned} \alpha &\equiv (-1 + 2 \cos \varphi) \sin(N\varphi) - \sin[(N-1)\varphi], \\ \beta &\equiv (-1 + 2 \cos \varphi) \cos(N\varphi) - \cos[(N-1)\varphi]. \end{aligned} \quad (3.63)$$

Non trivial solutions for A and B in the set of equations (3.62), require that the coefficient determinant D be zero. Using (3.63) one finds:

$$D = \beta \sin \varphi + \alpha(1 - \cos \varphi) = 2(-1 + \cos \varphi) \sin(N\varphi), \quad (3.64)$$

stating that the problem can be solved provided that the parameters φ be

$$\varphi_v = \frac{v\pi}{N}, \quad (3.65)$$

with v an integer. In conclusion, one has:

$$\omega(v) = \sqrt{\frac{4\kappa}{M}} \left| \sin \frac{\varphi_v}{2} \right|. \quad (3.66)$$

Concept

The vibrational excitations of a finite linear chain of N identical atoms are dispersed into a sinusoidal band, since they depend on an integer number $v = 1, \dots$, running up to the number of atoms N . Due to the harmonic approximation they are also independent.

For any fixed v one first calculates the normalization constant P appearing in the expression for the coordinates Q_n , by imposing the condition $P^2 \sum_{n=1}^N Q_n^2 = 1$. The detailed calculation is reported in Appendix 3.14, where it is found that $P^2 = 2[N(A^2 + B^2)]^{-1}$. Then, the first equation in (3.62) can be cast in the form

$$\begin{aligned} 0 &= -2A \sin\left(\frac{\varphi_v}{2}\right) \cos\left(\frac{\varphi_v}{2}\right) + 2B \sin^2\left(\frac{\varphi_v}{2}\right) \\ &= 2 \sin\left(\frac{\varphi_v}{2}\right) \left[-A \cos\left(\frac{\varphi_v}{2}\right) + B \sin\left(\frac{\varphi_v}{2}\right) \right]. \end{aligned} \quad (3.67)$$

Finally, the general solution to (3.59) is:

$$\begin{aligned} s_n(t) &= \sqrt{\frac{2}{N}} \sum_{v=1}^N a_v(t) \left[\frac{A_v}{\sqrt{A_v^2 + B_v^2}} \sin(n\varphi_v) + \frac{B_v}{\sqrt{A_v^2 + B_v^2}} \cos(n\varphi_v) \right] \\ &= \sqrt{\frac{2}{N}} \sum_{v=1}^N a_v(t) \left[\sin \frac{\varphi_v}{2} \sin(n\varphi_v) + \cos \frac{\varphi_v}{2} \cos(n\varphi_v) \right]. \end{aligned} \quad (3.68)$$

Here, A_v and B_v the coefficients of the displacements Q_n and

$$\sin \left(\frac{\varphi_v}{2} \right) = \frac{A_v}{\sqrt{A_v^2 + B_v^2}}, \quad \cos \left(\frac{\varphi_v}{2} \right) = \frac{B_v}{\sqrt{A_v^2 + B_v^2}}. \quad (3.69)$$

In (3.68), the normal variables $a_v(t)$ are

$$a_v(t) = p_v e^{i\omega(v)t} + q_v e^{-i\omega(v)t}, \quad (3.70)$$

with p_v and q_v fixed by the initial conditions.

3.5.3 Linear atomic chain with fixed edge atoms

In this case, inspection of the set of equations (3.59) immediately reveals that the situation where either one of the edge atoms is fixed, that is $s_1(t) = 0$ or $s_N(t) = 0$, implies only the trivial solution $s_n(t) = 0$ for $n = 1, \dots, N$. Thus, use of the boundary conditions $s_1(t) = 0$ and $s_N(t) = 0$ requires the introduction of two forces F_1 and F_N in the first and last equations of the set (3.59). The two forces must satisfy the conditions $F_1 + \kappa s_2 = 0$ and $F_N + \kappa s_{N-1} = 0$, so that the equations for the remaining atoms are not altered. The equations for $n = 2, \dots, N-1$ are once again sought for in the form $Q_n = A \sin(n\varphi) + B \cos(n\varphi)$. This results once again into relation (3.61). In addition, one must have

$$\begin{aligned} Q_1 &= A \sin \varphi + B \cos \varphi = 0, \\ Q_N &= A \sin(N\varphi) + B \cos(N\varphi) = 0. \end{aligned} \quad (3.71)$$

The condition of vanishing determinant yields $\sin[(N-1)\varphi] = 0$, and the frequencies are once again given by (3.61), though

$$\varphi_v = \frac{v\pi}{N-1}$$

instead. The normalization coefficient P can be calculated from the condition $P^2 \sum_{n=2}^{N-1} Q_n^2 = 1$ as in Appendix 3.14, yielding $P^2 = 2[(N-2)(A^2 + B^2)]^{-1}$. Collecting all the results together, the general solution is

$$s_n(t) = \sqrt{\frac{2}{N-2}} \sum_{v=1}^N a_v(t) \left[\frac{A_v}{\sqrt{A_v^2 + B_v^2}} \sin(n\varphi_v) + \frac{B_v}{\sqrt{A_v^2 + B_v^2}} \cos(n\varphi_v) \right]. \quad (3.72)$$

Using the first line in (3.71), this can be simplified into:

$$s_n(t) = \sqrt{\frac{2}{N-2}} \sum_{v=1}^N a_v(t) [\cos \phi_v \sin(n\phi_v) - \sin \phi_v \cos(n\phi_v)]. \quad (3.73)$$

Comparison of (3.68) with (3.73) makes clear that

Concept

Different boundary conditions determine different expressions for the allowed vibrational phases ϕ_v and thus normal mode frequencies.

In all cases these are proportional to $v\pi$ and inversely proportional to $N + \delta$, with δ depending on the boundary condition. In particular, $\delta = 0$ for free edge atoms and $\delta = -1$ for fixed edge atoms.

The limiting behavior of these solutions while the system evolves from a finite to a infinite chain with $N \rightarrow \infty$ can be studied along the same lines used to investigate the electronic structure in Chap. 2. Similar procedures and considerations hold, and are not repeated here, but one and the following exception.

A complication occurs indeed while treating vibrational modes: here, the Q_n are expressed as linear combinations of $\sin(n\phi)$ and $\cos(n\phi)$. This hurdle hands the opportunity of simplifying the above results. One can indeed consider a chain with $N + 2$ atoms instead, arranged from $x = 0$ to $x = (N + 1)a$ and pace a , with the condition of vanishing displacements of the edge atoms. This is precisely the same problem as that treated in the present section. Equation (3.61) yields the frequencies, with $\phi_v = v\pi/(N + 1)$. The hurdle is now knocked down since the displacements that satisfy the boundary conditions are now represented by the simplest expression $Q_n = A \sin(n\phi)$.

In conclusion, the following procedure can be conveniently followed to calculate vibrational frequency spectra and their corresponding eigenvalues:

Procedure

Step 1. Consider a system with $N + 2$ atoms with vanishing displacements for the edge atoms as boundary conditions.

Step 2. Write the equations of motion for the displacements $s_n(t)$, explicitly distinguishing those for the edge atoms.

Step 3. Seek solutions of the form $s_n(t) = Q_n e^{i\omega t}$ with $Q_n = A \sin(n\phi)$.

Step 4. Find the expression for ω^2 from the equations of motion for the in-chain atoms.

Step 5. Find the expression for φ , by imposing the condition of vanishing determinant for the two equations of motion referring to the edge atoms.

Step 6. Compute the normalization coefficient P as in Appendix 3.14.

Step 7. Wrap up all the information to write the resulting eigenvectors $s_n(t)$.

This procedure is being repeatedly applied to cases with increasing degrees of complexity.

3.5.4 Chain of $2N$ atoms with two alternating different species and first-neighbor interactions

One more complication is now introduced: atoms of two different species with masses M_1 and M_2 are here considered, that are staggered in the chain, alternating at the evenly spaced equilibrium positions. A convenient notation is introduced, where the positions of masses M_1 is labeled by, say, odd indexes and those of M_2 by even indexes. The problem of fixed edge atoms is here considered.

Therefore, the set of equations (3.59) can be adapted to the present situation, by keeping the edge atoms fixed at $x = 0$ and $x = (2N + 1)a$, as discussed at the end of the previous section. The equation for the left-edge atom with mass M_1 is obtained from the $n = 2$ equation in the set (3.59), with $s_0 = 0$. The equation of the right-edge atom with mass M_2 is similarly obtained by the second last equation in the set (3.59), labeled by index $2N$, and with $s_{2N+1} = 0$. According to this notation, the resulting new set of equations is:

$$\begin{aligned} M_1 \ddot{s}_1 &= \kappa(s_2 - 2s_1), & \text{for } n = 1, \\ M_1 \ddot{s}_{2n+1} &= \kappa(s_{2n+2} - s_{2n+1}) + \kappa(s_{2n} - s_{2n+1}), & \text{for } n = 1, 2, \dots, N-1, \\ M_2 \ddot{s}_{2n} &= \kappa(s_{2n-1} - s_{2n}) + \kappa(s_{2n+1} - s_{2n}), & \text{for } n = 1, 2, \dots, N-1, \\ M_2 \ddot{s}_{2N} &= \kappa(s_{2N-1} - 2s_{2N}), & \text{for } n = N. \end{aligned} \quad (3.74)$$

Setting as before $s_n = Q_n e^{-i\omega t}$, the homogeneous linear system is found

$$\begin{aligned} -\omega^2 M_1 Q_1 &= \kappa(Q_2 - 2Q_1), & \text{for } n = 1, \\ -\omega^2 M_1 Q_{2n+1} &= \kappa(Q_{2n+2} - Q_{2n+1}) + \kappa(Q_{2n} - Q_{2n+1}), & \text{for } n = 1, \dots, N-1, \\ -\omega^2 M_2 Q_{2n} &= \kappa(Q_{2n-1} - Q_{2n}) + \kappa(Q_{2n+1} - Q_{2n}), & \text{for } n = 1, \dots, N-1, \\ -\omega^2 M_2 Q_{2N} &= \kappa(Q_{2N-1} - 2Q_{2N}), & \text{for } n = N. \end{aligned} \quad (3.75)$$

Setting now

$$Q_{2n+1} = A \sin[(2n+1)\varphi_v],$$

$$Q_{2n} = B \sin [2n\phi_v]. \quad (3.76)$$

the second and third lines in (3.75) are satisfied provided that

$$\begin{aligned} (2\kappa - M_1 \omega^2) A - 2\kappa B \cos \phi_v &= 0, \\ -2\kappa A \cos \phi_v + (2\kappa - M_2 \omega^2) B &= 0. \end{aligned} \quad (3.77)$$

Solution of the equations above yield the frequencies to be:

$$\frac{\omega_v^2}{\kappa} = \frac{M_1 + M_2 \pm \sqrt{(M_1 + M_2)^2 - 4M_1 M_2 \sin^2 \phi_v}}{M_1 M_2}. \quad (3.78)$$

The first and fourth lines in (3.75) are instead satisfied provided that

$$\phi_v = \frac{v\pi}{2N+1},$$

precisely as anticipated in the previous paragraph. The integer values v now run from $v = 1, \dots, N$, since the set of $2N$ equations has been reverted back to a set of N identical pairs of equations like (3.77). Expression (3.78) states that when two species alternate in the chain, two phonon bands $\omega_s(v)$ occur with $s = 1$ and $s = 2$ labeling the solutions with negative and positive sign, respectively.

As to the eigenvectors, Appendix 3.14 details once again the calculation of normalization coefficients. The general solution to (3.75) is:

$$s_n(t) = \sum_{v=1}^N \sum_{s=1,2} Q_n(v, s) a_{v,s}(t), \quad (3.79)$$

with $a_{v,s}$ the normal modes variables,

$$Q_n(v, s) = \begin{cases} A_{v,s} \sin(n\phi_v), & n \text{ odd} \\ B_{v,s} \sin(n\phi_v), & n \text{ even.} \end{cases} \quad (3.80)$$

and finally

$$\begin{aligned} A_{v,s} &= \sqrt{\frac{4}{2N+1}} \left[1 + \left(\frac{2\kappa - M_1 \omega_s^2(v)}{2\kappa \cos \phi_v} \right)^2 \right]^{-\frac{1}{2}}, \\ B_{v,s} &= \frac{2\kappa - M_1 \omega_s^2(v)}{2\kappa \cos \phi_v} A_{v,s}. \end{aligned} \quad (3.81)$$

A few remarks are in order:

1. The band with $\omega_1(v)$ is an increasing function of v with a minimum at $v = 1$ and maximum at $v = N$.

Definition

A band with these characteristics is named acoustical band.

2. The band with $\omega_2(v)$ is a decreasing function of v with a maximum at $v = 1$ and minimum at $v = N$.

Definition

A band with these characteristics is named optical band.

3. In the limit $\varphi_v \ll 1$, the dispersions of the two bands are approximated by

$$\omega_1(v) \simeq \sqrt{\frac{2\kappa}{(M_1 + M_2)}} \frac{v\pi}{2N + 1}$$

and

$$\omega_2(v) \simeq \sqrt{\frac{2\kappa}{\mu}} \left(1 - \frac{M_1 M_2 v^2 \pi^2}{2(M_1 + M_2)^2 (2N + 1)^2} \right),$$

with μ the reduced mass of the two atoms. The frequency $\omega_1(v)$ is linear in v , with the angular coefficient increasing with decreasing N . The frequency $\omega_2(v)$ is quadratic in v , with a coefficient that increases while N decreases.

4. A gap opens up between the two bands, with size $\Delta\omega \equiv \omega_2(N) - \omega_1(N)$ given by

$$\Delta\omega = \frac{2\kappa}{M_1 M_2} \sqrt{(M_1 + M_2)^2 - 4M_1 M_2 \sin^2 \left(\frac{N}{2N + 1} \pi \right)}. \quad (3.82)$$

The minimum gap value is reached while $N \rightarrow \infty$ and the maximum at $N = 1$

5. The band width $\Delta \equiv \omega_2(1) - \omega_1(1)$, including the gap, is given by

$$\Delta = \frac{2\kappa}{M_1 M_2} \sqrt{(M_1 + M_2)^2 - 4M_1 M_2 \sin^2 \left(\frac{1}{2N + 1} \pi \right)}, \quad (3.83)$$

getting its maximum value while $N \rightarrow \infty$ and minimum value at $N = 1$.

In conclusion, the examples discussed so far imply that

Concept

The frequencies of normal vibrational modes depend on the number of atoms in the system, the dependence washing away while evolving from a finite towards an infinite system. Here, the frequencies become independent of the originally imposed boundary conditions.

The treatment of interactions extending to second and higher order neighbors proceeds along the same lines, though the complexity is so much higher that it is not considered here.

Quick Questions

Q8. Consider two chains with two same atoms with same mass and linked by same forces. Chain A has free ends, chain B has ends fixed. Which frequency of the mode with $v = 1$ is the higher?

Answer. For the mode $v = 1$, the phases are $\varphi_1 = \pi/N$ for chain A and $\varphi_1 = \pi/(N - 1)$ for chain B. Looking at the expression for the vibrational frequencies, the chain with free ends has the lower frequency.

Q9. You need to design a 1D atom-based device whose working concept requires flat vibrational modes. Discuss at least one obvious strategy to accomplish the task.

Answer. The bandwidth of vibrational modes for 1D atomic chains increases with increasing the number of atoms, being maximum in chains with length much larger than atom spacing. Thus, the most obvious strategy is to start from a chain made by the lowest number of atoms that is affordable.

3.6 Comments on systems with finite size

The content of Chap. 2 and of the present section are paralleled, since electronic and vibrational excitations have been determined starting from a finite chain composed first by one species and then by two species particles. Except for differences emerging from the different types of imposed boundary conditions, similar behaviors can generally be drawn. A few similarities are remarked in the present section.

One for all, the dispersive character of eigenfrequencies, basically originated by the lattice structure. Consider the situation with only one particle per cell. The same excitation structure emerges for both electron and phonon excitations: the states group themselves into bands, their width increasing up to a maximum finite value, while the system evolves into an infinite one with $N \rightarrow \infty$. In the situation with two species per cell, two bands originate that are separated at zone boundary, the corresponding gap getting smaller down to a minimum finite value, while the number of atoms increases towards the infinite system situation.

The question arises whether experiments exist which confirm these predictions. In Chap. 2 the case of a chain with gold atoms has been illustrated, that supports this view. Studies of vibrational states in finite systems are performed by using quantum nanostructures. Quantum lattices of GaAs/AlAs are especially suited to study the optical band frequencies of single GaAs or AlAs structures. the mass values of Ga

ad Al are significantly different, so that the frequencies of optical modes lie into distinguishable regions. To this aim, one layer of GaAs is usually sandwiched between two $\text{Al}_{1-x}\text{Ga}_x\text{As}$ layers. Under these circumstances, the vibrational states of the former material are represented by wavefunctions unable to penetrate the adjacent layers, but for selected values of x . As a consequence, the frequencies of optical modes are expected to increase while the layer thickness decreases. This prediction has been confirmed in the experiment [6].

The same type of quantum structures can be used to devise a second method aimed to test the dependence of either vibrational or electronic bands on the number of atoms and thus on finite-size effects. This time, the electronic valence and conduction bands of GaAs are considered. If the GaAs gap $E_g^{(1)}$ is smaller than the gap $E_g^{(2)}$ of $\text{Al}_{1-x}\text{Ga}_x\text{As}$, the nanostructure parameters can be tuned so that the wavefunction of an exciton be localized inside the GaAs layer. The energy released when the exciton decays is thus expected to depend on the thickness of GaAs layer: the thinner the layer, the larger the energy. A measurement of this energy, e.g. by optical means, infers the wanted information on the dependence of the excitation frequencies on the number of atoms, as performed in the experiment [15].

3.7 Vibrational modes of crystals

The normal variables have been introduced in Sec. 3.4 to solve the Schrödinger equation of interacting ions. Those results are here specialized to the case of a crystalline system. In this case, index J must embody a number of indexes, precisely the set \mathbf{R}_n , μ and i , that identify the cell

$$\mathbf{R}_n = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3,$$

the given atomic species μ in the base and the i -th component of the displacement. The equation of motion (3.48) thus reads:

$$M_\mu \ddot{s}_{m,\mu,i} = - \sum_{\underline{n}, \nu, l} \phi_{m,\mu,i}^{n,\nu,l} s_{n,\nu,l}, \quad (3.84)$$

where m and n indicate a tern of integer numbers identifying a lattice position. The coefficients $\phi_{m,\mu,i}^{n,\nu,l}$ have a physical meaning. In particular, $\phi_{m,\mu,i}^{n,\nu,l}$ is the force acted in direction i on an atom of species μ in the cell identified by \mathbf{R}_m . This force is originated by the unitary displacement in direction ℓ of an atom of species ν in the cell at \mathbf{R}_n . The quantity $-\sum_{n,\nu,l} \phi_{m,\mu,i}^{n,\nu,l}$ is instead the total force acted on the atom at \mathbf{R}_m of species μ in direction i , due to unitary displacements of all the crystal atoms, in direction ℓ . This force originates from a rigid shift of the whole crystal, so that the sum over all sites n and atomic species ν is to vanish:

$$\sum_{n,\nu} \phi_{m,\mu,i}^{n,\nu,l} = 0. \quad (3.85)$$

Equation (3.85) provides a first symmetry condition to be satisfied. A second condition arises by noticing that each translation labeled by a tern of integers h , is to leave the forces unchanged. Thus,

$$\phi_{m+h,\mu,i}^{n+h,v,l} = \phi_{m,\mu,i}^{n,v,l}, \quad \forall h,$$

and therefore

$$\phi_{m,\mu,i}^{n,v,l} = \phi_{m-n,\mu,i}^{0,v,l}. \quad (3.86)$$

In essence, whatever the complexity of the system might be, the structure of the equation of motion for the displacements (3.84) remains that of a harmonic oscillator, with constraints provided by general symmetry conditions, such as those dictated by zero-force theorem and by translational considerations. The conditions (3.85) and (3.86) are relevant in applications, as it is being discussed in the following, in a simple 1D geometry to begin with.

The above derivation can be coded into a Procedure as follows:

Procedure

Step 1. Begin with equations (3.84) along with the constraints (3.86).

Step 2. Set $s_m(t) \approx Q_m e^{-i\omega t}$.

Step 3. Insert the ansatz for $s_m(t)$ to derive equations for the $Q_m \approx e^{imka} Q_0$.

Step 4. Solve the resulting eigenvalue equation, to find eigenfrequencies and eigenvectors.

The procedure is now being applied to a number of cases, with increasing degree of complexity.

3.7.1 One dimensional crystal with one atom per cell

In this case indexes μ and i are immaterial. From (3.84) and (3.86) it follows that:

$$\ddot{s}_m = - \sum_n D_{m-n} s_n, \quad (3.87)$$

with $D_n = \phi_n/M$. Setting $s_m(t) = Q_m e^{-i\omega t}$, one has

$$\omega^2 Q_m = \sum_n D_{m-n} Q_n. \quad (3.88)$$

The crystal with N atoms under study is part of an infinite crystal, so that solutions are conveniently sought for, which satisfy periodic boundary conditions $Q_0 = Q_N$.

Indeed, consider the eigenvector from (3.88) represented through its components

$$\begin{pmatrix} Q_0 \\ \vdots \\ Q_{N-1} \end{pmatrix}, \quad (3.89)$$

and the corresponding eigenvalues along with it. Appendix 3.15 demonstrates that the eigenvector components Q_m are related to component Q_0 through the relation

$$Q_m = e^{imka} Q_0, \quad (3.90)$$

that is an alternative statement of Bloch theorem for vibrational states.

Equation (3.88) implies

$$\begin{aligned} \omega^2 Q_0 e^{imka} &= \sum_n D_{m-n} Q_0 e^{in ka}, \\ 0 &= \left(\omega^2 - \sum_p D_p e^{-ipka} \right) Q_0. \end{aligned} \quad (3.91)$$

Since $\phi_p = \phi_{-p}$ and mixed second derivatives are unchanged after exchange of coordinates in (3.47), the terms with $\sin(pka)$ disappear. As a consequence, one has

$$\omega^2 = \sum_p D_p \cos(pka). \quad (3.92)$$

Whenever $D_p \neq 0$, the elastic interaction acts as well between the p -th neighbors. As a general phenomenon, interactions between atoms are screened by the presence of electrons, as discussed in Chap. 4. One might thus conclude to good approximation that these interactions become negligible from a given order \bar{p} on. If this is the case, in (3.92) one can set $D_p = 0$ for each $|p| > \bar{p}$. The allowed values of k , that are those for which the boundary conditions $Q_0 = Q_N$ are satisfied, are $k = 2\pi n/(aN)$. From (3.92) the symmetries $\omega^2(k) = \omega^2(-k)$ and $\omega^2(k) = \omega^2(k + G)$ emerge. A similar symmetry holds for the eigenvectors, since from (3.90) they are seen to be left unchanged after rigid shifts by reciprocal lattice vectors $k \rightarrow k + G$. Thus, the k -values can be limited within the Brillouin zone, where N values of k result to be allowed, precisely the number that is needed to construct the general solution. Using these findings, the transformation (3.54) becomes

$$q_m(t) = \sum_k a_k(t) Q_0 e^{imka}, \quad (3.93)$$

with $k = 2\pi n/(aN)$ within the Brillouin zone.

Expression (3.92) yields the frequencies of the N normal modes as functions of wavevector \mathbf{k} within the Brillouin zone. The knowledge of the function $\omega(k)$ is hardly determined in general, without the values of the coefficients ϕ_p are first found. However, significant expressions can be obtained in the long wavelength

limit $k \rightarrow 0$. In this limit, one has

$$\omega^2 = \sum_p D_p \left[1 - \frac{(pka)^2}{2} \right]. \quad (3.94)$$

and using (3.85) one obtains

$$\omega = ck, \quad (3.95)$$

where $c = \sqrt{-(\sum_p D_p p^2 a^2)/2}$ and $c > 0$ because the original quadratic form is positive definite. The long-wavelength dispersion of normal modes is linear in k .

Examples

Consider the explicit case where the forces between ions only act between first neighbors. The equations of motion are

$$M\ddot{s}_m = \kappa(s_{m+1} - s_m) + \kappa(s_{m-1} - s_m). \quad (3.96)$$

Given the boundary conditions $Q_0 = Q_N$, the following dispersion relations are obtained:

$$\begin{aligned} \omega(k) &= \sqrt{\frac{4\kappa}{M}} \left| \sin\left(\frac{ka}{2}\right) \right|, \\ k &= \frac{2\pi n}{aN}. \end{aligned} \quad (3.97)$$

The left panel of Fig. 3.5 displays the spectrum of the normal modes represented by (3.97). This coincides with that found in (3.66) except for the different choice of the allowed values ϕ_v and k .

The above example enlightens the physical reasons originating the linear behavior of $\omega(k \rightarrow 0)$ in the long wavelength limit. In essence, in this limit (3.90) implies that

Concept

Atomic displacements of an increasingly larger number of neighboring cells are almost the same due to the short-range nature of the screened interactions between ions. Therefore, creating an excitation by e.g. displacing a far-away atom requires a vanishingly small energy, that is proportional to the excitation frequency.

Alternatively, due to (3.90) the discrete variables $s_n(t)$ can be substituted by the continuous $s(x, t)$. Equation (3.96) thus becomes

$$M \frac{\partial^2 s(x, t)}{\partial t^2} = \kappa a \frac{\partial s(x, t)}{\partial x} \Big|_{x=ma} - \kappa a \frac{\partial s(x, t)}{\partial x} \Big|_{x=(m-1)a} = \kappa a^2 \frac{\partial^2 s(x, t)}{\partial x^2}, \quad (3.98)$$

that is the equation for deformation waves in a continuous medium. The speed of sound in this medium is

$$c = \sqrt{\frac{\kappa a^2}{M}} = \sqrt{\frac{T}{\rho}}, \quad (3.99)$$

with $\rho = M/a$ the linear mass density and $T = \kappa a$ a tension. To all respects, the system behaves as if the atoms interacting with each other and vibrating about their equilibrium positions, were a continuous elastic medium with density ρ , each infinitesimal portion of it being subjected to a tension $T = \kappa a$. The quantity c is the phase velocity with which the deformation $s(x, t)$ propagates in the medium, that is the speed of sound-wave propagation inside the material. Now, in the case of continuous media the elastic modes have indeed been found to disperse according to a linear law in k , with a proportionality coefficient of the same form as (3.99). Therefore, the considered $k \rightarrow 0$ limit is the result of the continuum theory developed in Sec. 3.2.

In particular, the elastic constants c_{ij} and the speed of sound c are intimately related to each other. Deviations of $\omega(k)$ from linear behavior occur when the atomic displacements in neighboring cells cannot be considered to be almost the same. For example, (3.90) implies that at the Brillouin zone boundaries $k = \pm\pi/a$, one has $Q_1 = -Q_0$, $Q_2 = Q_0$, and $Q_3 = -Q_0$, corresponding to a different situation that cannot be mapped into the continuum limit. This is displayed in the left panel of Fig. 3.5, where the behavior of $\omega(k)$ represented by the solid lines is seen to markedly deviated from the linear law represented by dashed lines.

The dispersion relation $\omega = \omega(k)$ determines the group velocity of a wave packet via the relation

$$v_g = \frac{d\omega}{dk} = \sqrt{\frac{\kappa a^2}{M}} \cos \frac{ka}{2}, \quad (3.100)$$

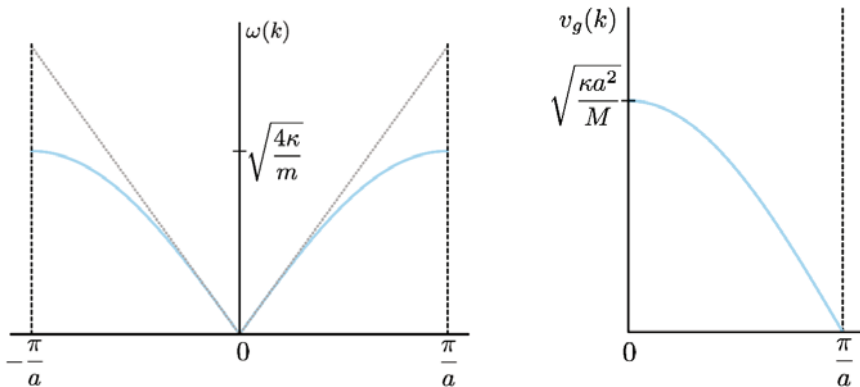


Fig. 3.5 Vibrational modes of 1D crystals with one atom per cell. Left panel: the frequency $\omega(k)$ is displayed within the Brillouin zone (solid lines). The deviation from linear behavior (dashed lines) is also made evident, as described in the text. Right panel: the group velocity of the vibrational wave

that is represented in the right panel of Fig. 3.5. The long wavelength limit $k \rightarrow 0$ has already been discussed and corresponds to the speed of sound $v_g = c$. The opposite, zone boundary, limit $k \rightarrow \pi/a$ yields $v_g(\pi/a) = 0$ and corresponds to a standing wave. The physical content of this finding is that the propagating vibrational wave undergoes a Bragg reflection whenever $\Delta k = k - (-k) = G \Leftrightarrow 2k = 2\pi/a$, that is, at the zone boundary. This is precisely what occurred with electron waves too, as discussed in Sec. 1.6.5.

Quick Questions

Q10. Consider a charged liquid, where long-range interactions predominate. What could you state on the behavior of the long-wavelength excitations?
Answer. In this case the displacements of an increasingly large number of neighboring particles could acquire significantly different sizes. Thus, exciting a long wavelength mode could require a finite amount of energy: the vibrational spectrum would have a gap at $k \rightarrow 0$. An example of such a situation is investigated in detail in Chap. 5.

Q11. Discuss how the interatomic interaction constants D_p could be experimentally determined.

Answer. From Eq. (3.92), it is easy to show that

$$D_p = \frac{a}{2\pi} \int_{-\pi/a}^{\pi/a} dk \omega^2(k) \cos(pka) .$$

Thus, a measurement of the dispersion relation $\omega(k)$ indirectly contains information on D_p . Notice that in metals long-range forces generally manifest up to several lattice spacing, so that the first-neighbor approximation used to derive (3.97), might be not valid.

3.7.2 One-dimensional crystals with two atoms per cell

Consider N cells evenly spaced by a , each containing two atoms, with only first-neighbors interactions. The equations of motion for the n -th cell are:

$$\begin{aligned} M_1 \ddot{s}_{n,1} &= \kappa(s_{n,2} - s_{n,1}) + \kappa(s_{n-1,2} - s_{n,1}) , \\ M_2 \ddot{s}_{n,2} &= \kappa(s_{n+1,1} - s_{n,2}) + \kappa(s_{n,1} - s_{n,2}) . \end{aligned} \quad (3.101)$$

Notice that in general the distance between atoms in a 1D crystal with two atoms per cell can be different from $a/2$. Therefore, it is convenient to somehow revert back to a situation with lattice spacing a and number of cells N . Setting $s_{n,i} = Q_{n,i} e^{-i\omega t}$, (3.101) turns into a set of homogeneous linear algebraic equations. Bloch theorem

(Appendix 3.15) holds, so that $Q_{n,i} = A_i e^{ikna}$. The algebraic system collapses into a set of identical systems, each with two equations. This is:

$$\begin{aligned} (-M_1 \omega^2 + 2\kappa)A_1 - \kappa(1 + e^{-ika})A_2 &= 0, \\ -\kappa(1 + e^{ika})A_1 + (2\kappa - M_2 \omega^2)A_2 &= 0. \end{aligned} \quad (3.102)$$

Periodic boundary conditions $Q_{0,i} = Q_{N,i}$ lead to the allowed k values $k = 2\pi n/(aN)$. The eigenfrequencies are then derived from (3.102), that is:

$$\begin{aligned} \frac{\omega^2}{\kappa} &= \frac{(M_1 + M_2) \pm \sqrt{(M_1 + M_2)^2 - 2M_1 M_2 [1 - \cos(ka)]}}{M_1 M_2} \\ &= \frac{(M_1 + M_2) \pm \sqrt{(M_1 + M_2)^2 - 4M_1 M_2 \sin^2(ka/2)}}{M_1 M_2}. \end{aligned} \quad (3.103)$$

In the long wavelength limit $k \rightarrow 0$, one gets

$$\begin{aligned} \omega_-(k) &= \sqrt{\frac{\kappa a^2}{2(M_1 + M_2)}} k = \sqrt{\frac{T}{\rho}} k, \\ \omega_+(k) &= \sqrt{2\kappa \frac{M_1 + M_2}{M_1 M_2}} \left[1 - \frac{1}{8} \frac{M_1 M_2}{(M_1 + M_2)^2} (ka)^2 \right], \end{aligned} \quad (3.104)$$

with $T = \kappa a/2$, $\rho = (M_1 + M_2)/a$. One immediately sees that two bands appear, one with $\omega_-(k)$ that is a linearly increasing function of k , and $\omega_+(k)$ that is a quadratic decreasing function of k starting from a constant value. These are the acoustical and optical bands, respectively. At the zone boundaries $k \rightarrow \pm\pi/a$ and assuming $M_1 > M_2$, one has instead

$$\begin{aligned} \omega_-(k) &= \sqrt{\frac{2\kappa}{M_1}} \left[1 - \frac{1}{8} \frac{M_1}{M_1 - M_2} (\pi - ka)^2 \right], \\ \omega_+(k) &= \sqrt{\frac{2\kappa}{M_2}} \left[1 + \frac{1}{8} \frac{M_2}{M_1 - M_2} (\pi - ka)^2 \right]. \end{aligned} \quad (3.105)$$

The nature of acoustical and optical bands is better understood if the amplitudes ratio $A_1/A_2 = \kappa(1 + e^{-ika})/(2\kappa - M_1 \omega^2)$ is calculated from (3.102) at both zone center and boundaries, that is for $k \rightarrow 0$ and $k \rightarrow \pm\pi/a$. In the two cases one finds

$$\begin{aligned} \left(\frac{A_1}{A_2} \right)_- &\rightarrow 1, \text{ acoustical band} \\ \left(\frac{A_1}{A_2} \right)_+ &\rightarrow \frac{2\kappa}{2\kappa - 2\kappa(M_1 + M_2)/M_2} = -\frac{M_2}{M_1}, \text{ optical band} \end{aligned} \quad (3.106)$$

at zone center, and

$$\left(\frac{A_1}{A_2}\right)_{-} \rightarrow \infty, \text{ acoustical band}$$

$$\left(\frac{A_1}{A_2}\right)_{+} \rightarrow 0, \text{ optical band ,}$$

at zone boundaries. At the zone center the acoustical band implies coincident displacement amplitudes A_1 and A_2 . This behavior reminds again the continuum limit, given that all the atoms have equal displacements independent of the species. The optical band implies displacements of the atoms in the cell that are characterized by opposite sign, so that their center of mass does not move. In this case the continuum approximation is not valid, since the two different atoms in the same cell have opposite-sign displacements and the same atoms in adjacent cells share the same displacement. This is why this band is named optical: when the two atomic species have opposite charge, the optical mode can be excited by the electric field carried by an electromagnetic wave with a frequency in the infrared region: the speed of light is indeed much larger than typical velocities of propagation for sound waves, and only the optical mode at $k \rightarrow 0$ can be selected. At the zone boundaries, the acoustical band corresponds to atoms of the same species in neighboring cells, that have opposite displacements, the other atoms in the cell remaining at rest. The optical band corresponds to a similar situation, but reversing the atom species. The two different situations are pictured in Fig. 3.6.

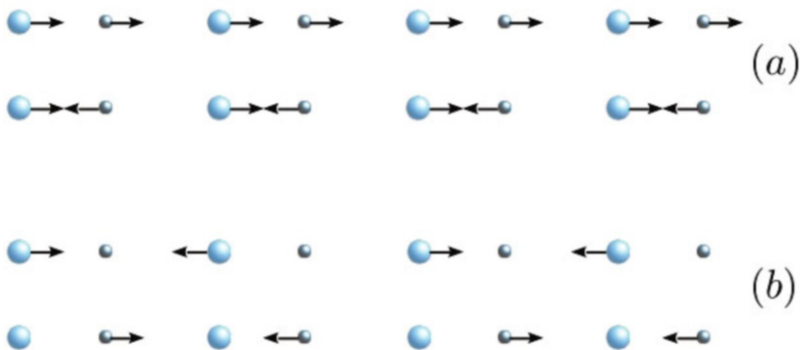


Fig. 3.6 Schematic picture of atomic motion (a) at the center and (b) at the zone boundaries, that correspond to the occurrence of optical (bottom lines) and acoustical (top lines) phonons

Expressions (3.104) and (3.105) provide both the gap $\Delta\omega$ between the two bands at $k = \pi/a$ and the total bandwidth Δ at $k = 0$, including the gap. Assuming $M_1 > M_2$, one has

$$\begin{aligned}\Delta\omega &= \omega_+\left(\frac{\pi}{a}\right) - \omega_-\left(\frac{\pi}{a}\right) = \sqrt{\frac{2\kappa}{M_2}} - \sqrt{\frac{2\kappa}{M_1}}, \\ \Delta &= \sqrt{\frac{2\kappa}{\mu}}\end{aligned}\quad (3.107)$$

with $\mu^{-1} = M_1^{-1} + M_2^{-1}$. The atom displacement in terms of normal mode functions $a_{k,s}(t)$ is

$$s_{m,\mu} = \sum_{k,s} a_{k,s}(t) e_s(k) e^{imka}, \text{ for } s = 1, 2. \quad (3.108)$$

Here, $e_s(k)$ represents the atom displacement in the reference cell at given k and with varying $s = 1, 2$. Fig. 3.7 shows the acoustical and optical bands of the 1D system with two atoms per cell.

Notice the mapping between the vibrational frequencies of the infinite 1D chain as in (3.103) and that of the finite chain as in (3.78), while taking into account that the boundary conditions are different in the two cases. The two results are in fact identical if one reminds that in the finite system $2N + 2$ represents the total number of atoms, while in the infinite system N represents the number of cells, each containing two atoms. Since $\varphi_v = v\pi/[2(N + 1/2)]$, the setting $ka/2 = v\pi/(2N + 1)$ realizes the mapping between the two expression, at least as long as $N \gg 1$.

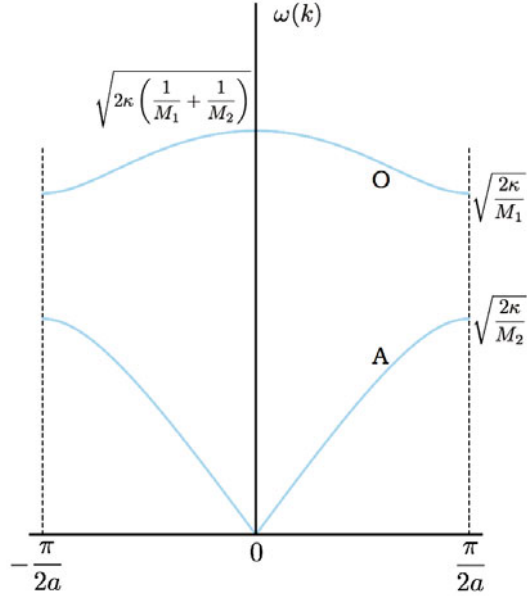


Fig. 3.7 Vibrational modes of 1D crystals with two atoms per cell. The frequency $\omega(k)$ of the acoustical (A) and optical (O) bands are displayed within the Brillouin zone. The gaps at Brillouin zone boundaries are evident

Quick Questions

Q12. The frequencies ν of vibration for a string with ends fixed are given

by the relation $\nu = \lambda \nu$, where λ is the wavelength of the stationary wave that sets in and ν the wave velocity. The fundamental vibration mode is the one that corresponds to no nodes along the length L of the string, thus to the condition $\lambda = L/2$: the longer the string, the lower the produced fundamental tone. Now, you need to build a piano, but your apartment is so small that not even a baby grand piano can fit. The issue is thus how to fabricate the strings for the lower tones. Once the tension and length of the string are fixed, how would you choose the material for the strings that must produce the lower tones?

Answer. Lower tones correspond to lower frequencies. Frequencies behave as $\approx \sqrt{1/\rho}$: thus materials with larger densities could be best suited. In fact, the low-tone strings in pianos are usually made thicker than high-tone strings, so to have a larger mass per unit length.

Q13. You are an orchestra director. Before starting the performance, you ask the wind instruments players to tune their instruments first. Is this weird or what?

Answer. Following the reasoning in Q9, the frequency of vibrational modes depends on density. For wind instruments, the vibrating string is air, and air density depends on temperature. Thus, the tuning of wind instruments is especially sensitive to temperature, at least with respect to string instruments for example. Since temperature cannot be nicely controlled inside the concert hall, it is therefore convenient to tune the wind instruments first and then tune the others on top of them.

3.7.3 Three-dimensional lattices

Equations (3.84) with the properties (3.86) are the basic tools to determine the vibrational spectra of whatever crystal, once periodic boundary conditions have been imposed on the crystal with N cells. The treatment of the general case with r atoms in the base is lengthy, and derived in detail for a simple case in Appendix 3.17. The present section is devoted to the case of one atom per cell. The result, already predictable with the experience accumulated so far, is that three bands appear, each of them with N vibrational acoustical modes with $\mathbf{k} \rightarrow 0$, their frequency linearly depending on the wavevector according to

$$\omega_s(\mathbf{k}) = c_s(\hat{\mathbf{k}})k, \quad (3.109)$$

with $s = 1, 2, 3$ and $c_s(\hat{\mathbf{k}})$ the phase velocity of sound waves along the direction of \mathbf{k} .

Fig. 3.8 Vibrational modes of 3D crystals with one atom per cell. The calculated vibrational spectrum of Al [7, 8], is displayed as solid lines along with the experimental data represented by dots. The three predicted acoustical branches are visible [9]. For details on theoretical and experimental setups, see [7, 8]. Reprinted with permission from [7, 8]. Copyright (1966) by the American Physical Society

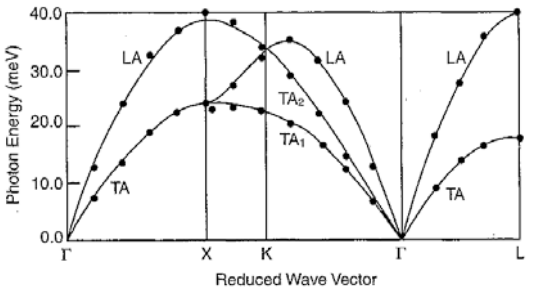
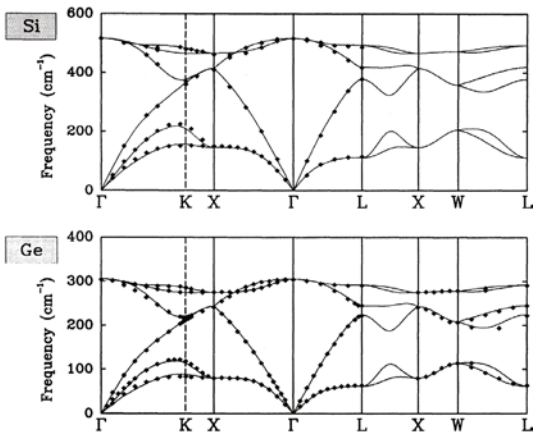


Fig. 3.9 Vibrational modes of 3D crystals with two atoms per cell. The calculated vibrational spectra of Si (top panel) and Ge (bottom panel) [10] are represented as solid lines along with the experimental data represented by dots. For details on theoretical and experimental setups, see [10]. Reprinted with permission from [10]. Copyright (1991) by the American Physical Society



Concept

In the more general case of a crystal with r atoms in the base, $3r$ vibrational bands appear, three of them being acoustical and the remaining $3r - 3$ optical. In the long wavelength acoustical modes, all the ions displace by the same amount independently of the species, while in the optical modes the ions move in each cell so to keep the center of mass be at rest. In each branch, N values of \mathbf{k} are allowed, as many as the number of cells.

Fig. 3.8 displays the vibrational bands of aluminum [7, 8], a material with one atom per cell. Fig. 3.9 displays instead the vibrational bands of silicon and germanium, that crystallize in an FCC with two atoms per cell [10].

Quick Questions

Q14. Is the concept of effective mass meaningful in the case of phonons?

Answer. Yes. Consider the critical points of the vibrational states energies $\hbar\omega_s(\mathbf{k})$ and perform series expansions around them up to second order. The effective inverse mass tensor comes out to have precisely the same form as in the electronic case.

Q15. You need to excite an acoustical vibrational mode first, and an optical mode afterward. Which strategies, or else which kind of excitation would you use in the two cases?

Answer. Considering the definition and conceptual meaning of acoustical and optical vibrations, an effective strategy would be to apply a pressure field for the acoustical mode and an electric dipolar field for the optical one.

3.8 Quantization of normal modes

3.8.1 Quantum nature of vibrational modes

The knowledge developed so far has been built up in a framework of classical equations of motion for the displacements, whatever discrete \mathbf{s}_m^μ or continuous $\mathbf{s}(\mathbf{r}, t)$ these displacements are. The description in terms of normal modes has been introduced, anticipating that it is suited to provide an easy extension to a quantum framework. Indeed, in Sec. 3.4 the procedure has been introduced to define the normal variables in classical mechanics, along with their relationship with the displacements. In Secs. 3.5 and 3.7 the same normal variables have been discussed for selected finite systems, leading to (3.68), (3.73), and (3.79), and for infinite periodic systems leading to (3.93), (3.108), and (3.207). In all the studied cases, the laws connecting normal variables and displacements result from unitary transformations. Thus, the quantization of the normal modes can be derived along the very same lines, starting from the Hamiltonians of independent harmonic oscillators, both for the finite and the infinite systems.

In the following, the result of the quantization procedure is given for the case of infinite crystals, where the dynamical variables are the atomic displacements \mathbf{s}_m^μ and the normal variables are $a_s(\mathbf{k})$. Begin with the unitary transformed Hamiltonian, that is:

$$\begin{aligned} H &= \sum_{\mathbf{k}, s} \frac{M}{2} [\dot{a}_{\mathbf{k}, s}^* \dot{a}_{\mathbf{k}, s} + \omega_s^2(\mathbf{k}) a_{\mathbf{k}, s}^* a_{\mathbf{k}, s}] \\ &= \sum_{\mathbf{k}, s} \left[\frac{1}{2M} p_{\mathbf{k}, s}^* p_{\mathbf{k}, s} + \frac{M}{2} \omega_s^2(\mathbf{k}) a_{\mathbf{k}, s}^* a_{\mathbf{k}, s} \right], \end{aligned} \quad (3.110)$$

where $p_{\mathbf{k},s} = M\dot{a}_{\mathbf{k},s}$ are the conjugate variables and the symmetry relations $p_{-\mathbf{k},s} = p_{\mathbf{k},s}^*$, $a_{-\mathbf{k},s} = a_{\mathbf{k},s}^*$, and $\omega_s(-\mathbf{k}) = \omega_s(\mathbf{k})$ hold. The system can be equivalently quantized starting either from the starting Hamiltonian or from (3.110). In any event, the energy eigenvalues are

$$E = \sum_{\mathbf{k},s} \hbar \omega_s(\mathbf{k}) \left(n_{\mathbf{k},s} + \frac{1}{2} \right). \quad (3.111)$$

The quantities $n_{\mathbf{k},s}$ are integers numbering the excitations in mode s with wavevector \mathbf{k} , the so called occupation numbers. The vibrational energy (3.111) is thus a sum of energies of single independent harmonic oscillator modes with energy $\hbar \omega_s(\mathbf{k})$ weighted in accord with their occupations $n_{\mathbf{k},s}$. These vibrational energy quanta are named phonons. They obey Bose-Einstein statistics. The vacuum mode with $n_{\mathbf{k},s} = 0$ anyway possesses the energy $E = \sum_{\mathbf{k},s} \hbar \omega_s(\mathbf{k})/2$. This so-called zero-point energy has no analogue in classical world. As to the eigenfunctions, these are coherently a product of wavefunctions of independent harmonic oscillators. In ultimate analysis, a vibrational state is defined when the number $n_{\mathbf{k},s}$ of excited energy quanta $\hbar \omega_s(\mathbf{k})$ is specified. Each quantum of energy is characterized by a well defined propagation wave vector \mathbf{k} and frequency $\omega = \omega_s(\mathbf{k})$. This is an example of quasiparticle.

Concept

A quasiparticle is a fictitious particle, with an assigned momentum $\hbar \mathbf{k}$ and energy $\hbar \omega$, useful to describe the excitations of a system. In the case of phonons, the quasiparticle describes a collective vibration of the atoms in the crystal. At variance with “true” particles, in the non-relativistic limit, a quasiparticle can be created and destroyed.

The quantum nature of vibrational modes, namely of the phonons, is evident in (i) the vanishing lattice contribution to thermal capacitance in the $T \rightarrow 0$ limit, and (ii) the discrete structure of inelastic diffusion spectra of X rays or neutrons from the lattice, showing that the lattice releases or absorbs energy in discrete quanta $\hbar \omega$. These facts are analyzed in Secs. 3.9 and 3.12.

Appendix 2.14 introduces the quite useful representation of harmonic oscillator states in terms of occupation numbers. Using those results, the creation $b_{\mathbf{k},s}^\dagger$ and destruction $b_{\mathbf{k},s}$ operators are conveniently introduced according to the following relations:

$$\begin{aligned} b_{\mathbf{k},s}^\dagger &= \sqrt{\frac{M\omega_s(\mathbf{k})}{2\hbar}} a_{\mathbf{k},s}^* - i\sqrt{\frac{1}{2\hbar M\omega_s(\mathbf{k})}} p_{\mathbf{k},s}, \\ b_{\mathbf{k},s} &= \sqrt{\frac{M\omega_s(\mathbf{k})}{2\hbar}} a_{\mathbf{k},s} + i\sqrt{\frac{1}{2\hbar M\omega_s(\mathbf{k})}} p_{\mathbf{k},s}^*. \end{aligned} \quad (3.112)$$

In the other way around:

$$\begin{aligned} a_{\mathbf{k},s} &= \sqrt{\frac{\hbar}{2M\omega_s(\mathbf{k})}} \left(b_{-\mathbf{k},s}^\dagger + b_{\mathbf{k},s} \right), \\ p_{\mathbf{k},s} &= -i\sqrt{\frac{\hbar M\omega_s(\mathbf{k})}{2}} \left(b_{-\mathbf{k},s} - b_{\mathbf{k},s}^\dagger \right). \end{aligned} \quad (3.113)$$

The Hamiltonian can therefore be expressed as

$$H = \sum_{\mathbf{k},s} \hbar\omega_s(\mathbf{k}) \left(b_{\mathbf{k},s}^\dagger b_{\mathbf{k},s} + \frac{1}{2} \right), \quad (3.114)$$

that is a weighted sum of operators $b_{\mathbf{k},s}^\dagger b_{\mathbf{k},s}$, acting to count the number of phonons with wavevector \mathbf{k} in mode s . Indeed, the creation and destruction operators are defined by the way in which they act on the ground state $|0\rangle$ with zero phonons and on the excited states $|n_{\mathbf{k},s}\rangle$ with $n_{\mathbf{k},s}$ phonons. That is:

$$\begin{aligned} b_{\mathbf{k},s}^\dagger |n_{\mathbf{k},s}\rangle &= \sqrt{n_{\mathbf{k},s} + 1} |n_{\mathbf{k},s} + 1\rangle, \\ b_{\mathbf{k},s} |n_{\mathbf{k},s}\rangle &= \sqrt{n_{\mathbf{k},s} - 1} |n_{\mathbf{k},s} - 1\rangle, \\ b_{\mathbf{k},s} |0\rangle &= 0, \\ b_{\mathbf{k},s}^\dagger b_{\mathbf{k},s} |n_{\mathbf{k},s}\rangle &= n_{\mathbf{k},s} |n_{\mathbf{k},s}\rangle. \end{aligned} \quad (3.115)$$

In essence, $b_{\mathbf{k},s}^\dagger$ ($b_{\mathbf{k},s}$) creates (destroys) one phonon on the state with $n_{\mathbf{k},s}$, resulting into a state with $n_{\mathbf{k},s} + 1$ ($n_{\mathbf{k},s} - 1$) phonons.

As to the displacements $\mathbf{s}_{m,\mu}$ of the μ species in cell m , these are represented by:

$$\mathbf{s}_{m,\mu} = \sum_{\mathbf{k},s} \sqrt{\frac{\hbar}{2M_\mu\omega_s(\mathbf{k})}} \left(b_{-\mathbf{k},s}^\dagger + b_{\mathbf{k},s} \right) \mathbf{e}_\mu(\mathbf{k},s) e^{i\mathbf{k}\cdot\mathbf{R}_m}, \quad (3.116)$$

namely a weighted superposition of plane waves in the lattice coordinates \mathbf{R}_m with wavevector \mathbf{k} and directed along the \mathbf{k} dependent polarization vector $\mathbf{e}_\mu(\mathbf{k},s)$ of the given mode s . Each plane wave is weighted by the typical length of the harmonic oscillator $\sqrt{\hbar/(2M_\mu\omega_s(\mathbf{k}))}$ for the corresponding mode. The displacements correspond to an excitation where a phonon is either destroyed or created, and thus the quantum operator must involve both b^\dagger and b operators. Since these excitations correspond to fluctuations of the density, the quantum operator creating them has to be symmetric and thus of the type $b_{-\mathbf{k}}^\dagger + b_{\mathbf{k}}$ with the $-\mathbf{k}$ and \mathbf{k} structure ensuring momentum conservation. Along the same lines, the corresponding momenta are represented by:

$$\mathbf{p}_{m,\mu} = i \sum_{\mathbf{k},s} \sqrt{\frac{\hbar M_\mu\omega_s(\mathbf{k})}{2}} \left(b_{-\mathbf{k},s}^\dagger - b_{\mathbf{k},s} \right) \mathbf{e}_\mu(\mathbf{k},s) e^{i\mathbf{k}\cdot\mathbf{R}_m}, \quad (3.117)$$

involving the antisymmetric, momentum-like combination $b_{-\mathbf{k},s}^\dagger - b_{\mathbf{k},s}$ and with the imaginary unit ensuring the hermitian property.

The rest of the treatment closely follows the classical case. The knowledge of phonon eigenfunctions and eigenvalues or energy spectra begins with the knowledge of the coefficients $\phi_{h,\mu,i}^{0,v,l}$ introduced in (3.47), that shape the potential in harmonic approximation and thus enter eigenvalues and eigenfunctions. The coefficients $\phi_{h,\mu,i}^{0,v,l}$ though, are the result of modeling and are not known a priori, so that their values are to be conveniently either computed from first principles, or fitted from experimental data on the spectra $\omega_s(\mathbf{k})$ or else on the elastic properties of materials.

3.8.2 Density of vibrational and electronic states

As it appears evident from the treatment above, all the calculations on phonon excitations involve weighted sums of various types of single-particle quantities, where the number of vibrational normal modes and the wavevector \mathbf{k} enter. Similar circumstances occur while determining electronic excitations. Here, the concept of density of states comes to help:

Concept

The density of vibrational or electronic states determines the weight with which a given mode contributes to build up the average quantity that is observable.

Thus,

Definition

The density of vibrational or electronic states is the number of normal modes living in the unit interval of frequencies or energies.

In essence, the concept of density of states conveniently converts sums over n and \mathbf{k} into sums or else integrals over frequency or energy.

Consider first a system having a spectrum of discrete normal modes frequencies. In this simple case, the density of normal modes $g(\omega)$ is simply given by:

$$g(\omega) = \sum_i g_i \delta(\omega - \omega_i), \quad (3.118)$$

where g_i counts the possible degeneracy of the normal mode with frequency ω_i . The number of modes lying in the interval between ω_a and ω_b is

$$G = \int_{\omega_a}^{\omega_b} g(\omega) d\omega = \sum_{\omega_a < \omega_i < \omega_b} g_i \quad (3.119)$$

Such expressions are valid for both vibrational and electronic states. The actual calculation of phonon and electron density of states in different dimensions is now in order.

3.8.2.1 Density of phononic normal modes in a crystal and critical points

3D case. In the case of periodic systems the bands of normal modes have been labeled by discrete indexes s and wavevector \mathbf{k} within the first Brillouin zone:

$$\mathbf{k} = \frac{n_1}{N_1} \mathbf{b}_1 + \frac{n_2}{N_2} \mathbf{b}_2 + \frac{n_3}{N_3} \mathbf{b}_3,$$

in terms of the integers n_i with $i = 1, 2, 3$. Here, $N_1 N_2 N_3 = N$ is the total number of direct lattice cells, where periodicity boundary conditions on the displacements have been imposed. It has been found that the frequencies of normal modes satisfy the conditions $\omega_s(\mathbf{k}) = \omega_s(-\mathbf{k})$ and $\omega_s(\mathbf{k}) = \omega_s(\mathbf{k} + \mathbf{G})$, with \mathbf{G} an arbitrary wavevector of the reciprocal lattice. Now,

Concept

Periodicity of $\omega_s(\mathbf{k})$ on the whole reciprocal lattice implies that in the first BZ $\omega_s(\mathbf{k})$ have at least one minimum, say M_0 , two saddle points, say M_1 and M_2 , and one maximum M_3 .

These are known as van Hove critical points. The emergence of the critical points, occurring also in the case of electronic bands, is clarified in the following by means of the density-of-state concept.

Consider the volume associated to each mode in the reciprocal lattice. This is $V_0 = V_r/N$ in terms of the Brillouin zone volume V_r and of the number of cells N in direct lattice. The number dN of frequencies lying in a volume $d\mathbf{k}$ of the BZ is thus

$$dN = \frac{d\mathbf{k}}{V_0} = \frac{N}{V_r} d\mathbf{k} = \frac{NV_c}{(2\pi)^3} d\mathbf{k} = \frac{V}{(2\pi)^3} d\mathbf{k}, \quad (3.120)$$

with V_c the volume of the elementary cell in direct lattice and V the total system volume. Consider now two surfaces on which the frequency of the phonon branch $\omega_s(\mathbf{k})$ is constant, say $\omega_s(\mathbf{k}) = \omega$ and $\omega_s(\mathbf{k}) = \omega + d\omega$. The two surfaces could in principle be inside the BZ or intersect its boundary planes. Be dk_\perp the wavevector perpendicular to the surfaces $\omega_s(\mathbf{k}) = \omega$ at \mathbf{k} , so that $d\mathbf{k} = dS dk_\perp$ with dS the area of the corresponding portion of surface at constant frequency. The number of states in volume $d\mathbf{k}$ identified by the two surfaces above is

$$\begin{aligned} dN_s &= \frac{V}{(2\pi)^3} d\mathbf{k} = \frac{V}{(2\pi)^3} \int_{\omega_s(\mathbf{k})=\omega} dS dk_\perp \\ &= \frac{V}{(2\pi)^3} \int_{\omega_s(\mathbf{k})=\omega} \frac{dS}{|\nabla \omega_s(\mathbf{k})|} d\omega. \end{aligned} \quad (3.121)$$

The second equality in (3.121) embodies the Jacobian $d\omega = |\nabla\omega_s(\mathbf{k})|dk_\perp$, transforming an integral over \mathbf{k} vectors into an integral over energies or else frequencies. Thus, the density of vibrational modes in branch $\omega_s(\mathbf{k})$ is:

$$g_s(\omega) = \frac{V}{(2\pi)^3} \int_{\omega_s(\mathbf{k})=\omega} \frac{dS}{|\nabla\omega_s(\mathbf{k})|}. \quad (3.122)$$

The explicit calculation of the density of states (3.122) is in general as hard to be executed as much complicated is the form of the surface with constant energy. The calculation may result simplified in selected cases. The important issue is the determination of the critical points, that are characterized by a vanishing gradient $\nabla\omega_s(\mathbf{k}) = 0$. In the vicinity of a critical point, the Taylor expansion of $\omega_s(\mathbf{k})$ up to second order is a quadratic form. This can be written in canonical form

$$\omega_s(\mathbf{k}) = \omega_s(\mathbf{k}_0) + \alpha_x(k_x - k_{0x})^2 + \alpha_y(k_y - k_{0y})^2 + \alpha_z(k_z - k_{0z})^2, \quad (3.123)$$

in terms of the expansion coefficients α_x , α_y , and α_z . A point M_0 is characterized by positive coefficients α_i , with $i = x, y, z$, the quadratic form being thus positive definite. A point M_3 is characterized by negative coefficients, the quadratic form being negative definite. A point M_1 by two positive and one negative coefficients, while M_2 by one positive and two negative ones, the quadratic form being of undefined sign. The integration domain of (3.122) is thus represented by either ellipsoids for M_0 and M_3 or hyperboloids for M_1 and M_2 . The integral can be executed analytically and yields:

1. Point M_0 :

$$g_s(\omega) = \begin{cases} 0 & \text{for } \omega < \omega_0, \\ \frac{V}{4\pi^2} \frac{1}{\sqrt{\alpha_x\alpha_y\alpha_z}} \sqrt{\omega - \omega_0} & \text{for } \omega > \omega_0. \end{cases} \quad (3.124)$$

2. Point M_1 :

$$g_s(\omega) = \begin{cases} \frac{V}{8\pi^2} \frac{1}{\sqrt{|\alpha_x\alpha_y\alpha_z|}} (R - \sqrt{\omega_0 - \omega}) & \text{for } \omega < \omega_0, \\ \frac{V}{8\pi^2} \frac{1}{\sqrt{|\alpha_x\alpha_y\alpha_z|}} R \left(1 - \frac{\omega - \omega_0}{2R^2}\right) & \text{for } \omega > \omega_0, \end{cases} \quad (3.125)$$

with R the value of the density of states at the critical point.

3. Point M_2 :

$$g_s(\omega) = \begin{cases} \frac{V}{8\pi^2} \frac{1}{\sqrt{|\alpha_x \alpha_y \alpha_z|}} R \left(1 - \frac{\omega_0 - \omega}{2R^2}\right) & \text{for } \omega < \omega_0, \\ \frac{V}{8\pi^2} \frac{1}{\sqrt{|\alpha_x \alpha_y \alpha_z|}} (R - \sqrt{\omega - \omega_0}) & \text{for } \omega > \omega_0, \end{cases} \quad (3.126)$$

with R as above.

4. Point M_3 :

$$g_s(\omega) = \begin{cases} \frac{V}{4\pi^2} \frac{1}{\sqrt{|\alpha_x \alpha_y \alpha_z|}} \sqrt{\omega_0 - \omega} & \text{for } \omega < \omega_0, \\ 0 & \text{for } \omega > \omega_0. \end{cases} \quad (3.127)$$

Evidently, in all cases the occurrence of critical points manifests through different algebraic behaviors of the density of states, according to whether the point is approached from values of ω that are higher or lower than ω_0 . Critical points in three dimensions are also characterized by the continuity of $g_s(\omega \rightarrow \omega_0^\pm)$.

In essence, the following procedure is to be used to calculate the phonon density of states:

Procedure

Step 1. Determine the geometrical expression for the surface S with constant energy.

Step 2. Extract an expression for $\omega_s(\mathbf{k})$ by first principles or approximated calculations, or else by experimental data.

Step 3. Compute the surface integral (3.122).

Step 4. Analyze the limiting behavior of the resulting density of states, while the frequency approaches from above and from below the critical points, defined as the points for which $\nabla \omega_s(\mathbf{k}) = 0$.

The procedure is now applied to the example of acoustical phonons in the vicinity of $k = 0$.

Examples

In the case of long-wavelength acoustical phonons, it turns out that the constant-frequency surface is a sphere, whose radius depends on the direction of \mathbf{k} . Since indeed $k = \omega/f_s(\hat{\mathbf{k}})$, one has $dS = [\omega^2/f_s(\hat{\mathbf{k}})^2] d\Omega$ in terms of the infinitesimal solid angle $d\Omega$. Thus:

$$g_s(\omega) = \frac{V}{(2\pi)^3} \omega^2 \int \frac{d\Omega}{f_s(\hat{\mathbf{k}})^3} = \frac{V}{(2\pi)^3} \frac{4\pi\omega^2}{c_s^3}, \quad (3.128)$$

that is quadratic in the frequency ω . Here, the angular average $1/c_s^3$ of the cubed inverse speed of sound has been introduced. Therefore:

$$g_{ac}(\omega) = \frac{V}{(2\pi)^3} 4\pi\omega^2 \left(\frac{1}{c_1^3} + \frac{1}{c_2^3} + \frac{1}{c_3^3} \right).$$

Under special symmetry conditions, the latter can be expressed as

$$g_{ac}(\omega) = \frac{V}{(2\pi)^3} 4\pi\omega^2 \left(\frac{2}{c_t^3} + \frac{1}{c_l^3} \right), \quad (3.129)$$

in terms of the speeds of sound for transverse c_t and longitudinal c_l oscillations.

2D case. The calculation in a 2D system proceeds along the same lines as in the 3D case. One obtains:

$$g_s(\omega) \propto \begin{cases} \theta(\omega - \omega_0), & \text{at } M_0 \\ \ln|\omega - \omega_0|^{-1} & \text{at saddle points } M_1 \text{ and } M_2, \\ \theta(\omega_0 - \omega) & \text{at } M_3. \end{cases} \quad (3.130)$$

Two remarks are in order. First, the two saddle points have the same behavior originated by equal resulting coefficients. Second, the occurrence of critical points in two dimensions is manifested by a discontinuous behavior of $g_s(\omega \rightarrow \omega_0^\pm)$ for points M_0 and M_3 , and by a logarithmic divergence at M_1 and M_2 .

Examples

In the case of acoustical phonons in the vicinity of $k = 0$ one finds

$$g_{ac}(\omega) = \frac{S}{2\pi} \omega \left(\frac{1}{c_1^2} + \frac{1}{c_2^2} \right) \quad (3.131)$$

that is linear in ω , with c_s the speed of sound in the s -th direction, $s = 1, 2$ and S the system surface.

1D case. In this case, for M_0 and M_3 one has:

$$g_s(\omega) \propto \frac{1}{\sqrt{|\omega - \omega_0|}}, \quad (3.132)$$

showing a square-root divergence.

Examples

For the acoustical branch in the long-wavelength limit $k \rightarrow 0$ one has:

$$g_{ac}(\omega) = \frac{L}{2\pi} \frac{1}{c}, \quad (3.133)$$

that is a constant, in terms of the speed of sound c and the system length L .

In essence,

Concept

The type of critical points and the dimensionality of the system under study are reflected by characteristic and distinguished behaviors of the density of states with respect to frequency ω .

Reverting back to expression (3.97), the results (3.129) and (3.132) can be checked and verified. By explicit calculation of the density of states and using the definitions along with (3.97), one finds

$$g(\omega) = \frac{2}{\pi} N \frac{1}{\sqrt{\omega_M^2 - \omega^2}}. \quad (3.134)$$

Here, the band maximum $\omega_M \equiv \sqrt{4\kappa/M}$ has been defined. In the vicinity of $\omega = 0$, this yields $g(\omega) = 2N/(\pi\omega_M)$ that is constant as in the formula (3.133), while in the vicinity of $\omega = \omega_M$ it is found the form (3.132), as aspected.

3.8.2.2 Density of electronic modes in a crystal

The introduction of the concept of Density of States (DOS) for phonons makes easier the introduction of a similar and useful concept for electron states. This completes the useful tools to investigate electronic excitations and initiated in Chap. 2. The energy bands of electronic states have been seen to satisfy the properties $E_n(\mathbf{k}) = E_n(-\mathbf{k})$ and $E_n(\mathbf{k}) = E_n(\mathbf{k} + \mathbf{G})$ for each vector \mathbf{G} of the reciprocal lattice. Thus, the results obtained in the case of phonons can be extended to electrons, provided a factor of 2 is introduced to take into account the spin degeneracy. In detail, the frequencies ω are to be replaced by energies E , and the constants α_i by $\hbar^2/(2m_i^*)$, with m_i^* being the value of the i -th component of the principal effective mass tensor:

$$(m^*)_{i,j}^{-1} = \frac{1}{\hbar^2} \frac{\partial^2 E_n(\mathbf{k})}{\partial k_i \partial k_j}. \quad (3.135)$$

This procedure can be used in all cases, whether 1D, 2D, or 3D, with the caveat that in general no equivalent of acoustic bands exist for electron systems in the $k \rightarrow 0$ limit, they being charged.

An exception is graphene, whose band structure in the vicinity of the Fermi level is

$$E(k) = \epsilon_f \pm \frac{3}{2}a_0|\gamma|k, \quad (3.136)$$

with $v = \frac{3}{2}a_0\gamma$, as calculated in Chap. 2 in terms of the 2D wavevector \mathbf{k} . The situation here is similar to the case of acoustical phonons in 2D, and the density of states around the Fermi level indeed turns out to be:

$$g(E) = \frac{S}{2\pi v^2} |E - \epsilon_f|. \quad (3.137)$$

Quick Questions

Q16. Consider the frequency-to-wavevector relation (3.97) in the long-wavelength limit $k \rightarrow 0$. What is the value of the effective mass of the excitation?

Answer. As the quadratic term in the Taylor expansion of $\omega(k)$ is missing, the absolute value of effective mass turns out to be exceedingly large, tending to infinity.

Q17. Discuss at least one especially useful information that can be extracted from a measurement of $g_s(\omega)$.

Answer. The dimensionality of the system for example, but also the type of critical point.

3.9 Thermodynamics of vibrational states: harmonic effects

Analysis of specific heat data infer important information on excitations in crystals. Their study is useful to understand and predict thermodynamical behavior of systems.

Vibrational energy and specific heats in crystals. The relationship between specific heat and vibrational excitations in crystals is explored in the following.

The crystal heat capacity at constant volume is defined as

$$C_V = \left(\frac{\partial U}{\partial T} \right)_V, \quad (3.138)$$

where U is the total internal energy. The heat capacity measures the capability of a system to change its internal energy as a consequence of a temperature variation.

The experimental measurements on a wide class of materials bring out two main findings:

- Close to room temperature, almost all crystals show a constant heat capacity. The value turns out to be $3N_{at}k_b$ independent of the material, with N_{at} the number of atoms in the sample and k_b the Boltzmann's constant. For one mole of atoms, $N_{at} \simeq 6.03 \cdot 10^{23}$, the molar specific heat is $c_V \simeq 6 \text{ cal mole}^{-1} \text{ }^\circ\text{K}^{-1}$. This is usually referred to as the Dulong-Petit law.
- As $T \rightarrow 0$, the heat capacity approaches zero with a T^3 or a T power law, depending on whether the material is an insulator or a metal, respectively. A quantum approach is needed to explain these results.

As an example, the dependence of the measured heat capacity of silicon and germanium on temperature is reported in Fig. 3.10.

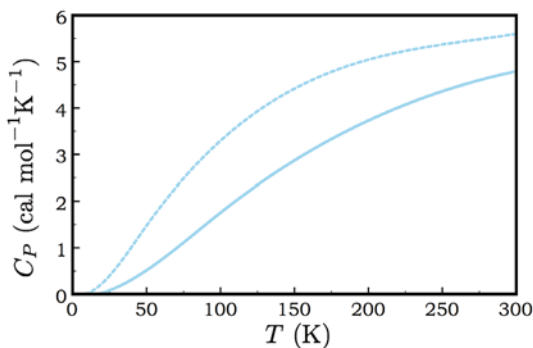


Fig. 3.10 Heat capacity vs. temperature. Solid curve: Silicon. Dashed curve: Germanium [11]

Appendix 3.16 details the calculations of free energy, entropy, and total vibrational energy in harmonic approximation. In this approximation, the frequencies of normal modes are seen to be independent of volume. To the aim of the present section, the relevant relation is the total energy in terms of the vibrational frequencies, as a function of temperature. This is:

$$U = U_0 + \sum_{\mathbf{k},s} \frac{\hbar\omega_s(\mathbf{k})}{2} + \sum_{\mathbf{k},s} \frac{\hbar\omega_s(\mathbf{k})}{e^{\hbar\omega_s(\mathbf{k})\beta} - 1}, \quad (3.139)$$

with $\beta = (k_b T)^{-1}$. The first two terms in (3.139) embody information on the structure of the material, U_0 being an overall constant, and do not depend on temperature. Indeed, the physical meaning of the last term is better caught after casting it as $\sum_{\mathbf{k},s} \hbar\omega_s(\mathbf{k}) n_s(\mathbf{k})$, with $n_s(\mathbf{k}) = [e^{\hbar\omega_s(\mathbf{k})\beta} - 1]^{-1}$ the average number of phonons with wavevector \mathbf{k} in band s , as resulting from the appropriate Bose-Einstein quantum statistics. The last term thus embodies the vibrational modes contribution to the specific heat.

Concept

In harmonic approximation, the vibrational modes contribution to specific heat is the sum of the vibrational energies $\hbar\omega_s(\mathbf{k})$ of the independent oscillator states, weighted by their average occupation at the given temperature, according to Bose-Einstein quantum statistics.

The finding that vibrational frequencies do not depend on volume within the harmonic approximation has relevant consequences on the description of the system response to thermodynamic perturbations. In crystals in fact, thermal expansion manifests as a displacement of the equilibrium position of atoms: the force driving the atoms out of equilibrium leads to a first-order, linear term in the potential energy expansion around equilibrium. Now, Problem 3.10 studies the simple case of a quadratic form, as it results from a second-order expansion from a non-stationary initial point, including a linear term in the expansion of harmonic potential energy. In this problem, it is seen that a reduction of this form to a purely quadratic one leaves the second-order coefficients unchanged. Given that the frequencies are also unchanged, applications of this result to a harmonic crystal lead to vanishing pressure $P = -(\partial U_1/\partial V)_T = 0$ and compressibility B , where $U_1 = \sum_{\mathbf{k},s} \hbar\omega_s(\mathbf{k})n_s(\mathbf{k})$. A similar reasoning can be elaborated for other characteristic responses of the system to various types of temperature changes. Thus,

Concept

Finite response of a crystal to mechanical or thermal perturbations, as it is measured by compressibility or thermal-expansion and conductivity coefficients, is embodied in the constant term U_0 analyzed in Chap. 1 and representing the cohesive-energy term, and by anharmonic terms beyond second-order.

This last concept is detailed in Sec. 3.10. In general, determination of the system response to thermal driving is conveniently performed by the following procedure:

Procedure

Step 1. Begin with the expression for the average energy (3.139), and turn the sums $\sum_{\mathbf{k},s} \dots$ over \mathbf{k} and s into the integral $\int \dots n_s(\omega, T) g_s(\omega) d\omega$, involving the appropriate phonon density of states $g_s(\omega)$ and phonon occupation numbers $n_s(\omega, T)$ at the given frequency ω and temperature T .

Step 2. Define the type of thermal driving: this can be a simple overall change of temperature value, or a local gradient of temperature, ...

Step 3. Define the observable that is expected to change after the thermal driving in terms of the average energy U , and expand it in the regime of small perturbations, i.e. linear in the driving. The observable can be the average energy itself (leading to

the determination of specific heat), or a volume change (leading to the thermal expansion coefficient), or an energy current transport (leading to thermal conductivity coefficient), and so on.

Step 4. Explicitly carry out the calculation and extract the wanted response coefficient.

In the following the procedure is applied to a number of examples under selected approximations. In particular, explicit calculation of the specific heat within harmonic approximation may proceed analytically in two opposite limiting behaviors.

Examples

In the high-temperature, classical limit $k_b T \gg \hbar \omega_M$, where ω_M is the largest phonon frequency, one may approximate (3.139) with

$$U = \sum_{\mathbf{k},s} \frac{\hbar \omega_s(\mathbf{k})}{e^{\hbar \omega_s(\mathbf{k})\beta} - 1} \simeq \sum_{\mathbf{k},s} \frac{\hbar \omega_s(\mathbf{k})}{\hbar \omega_s(\mathbf{k})\beta} = 3rNk_b T, \quad (3.140)$$

where the $3r$ phonon bands have been accounted for. Calculation of the corresponding specific heat from its definition leads to the following result for the specific heat per particle

$$c_V = \frac{1}{Nr} \left(\frac{\partial U}{\partial T} \right)_V = 3k_b. \quad (3.141)$$

The result above represents the Dulong and Petit law and can be easily shown to be a consequence of the energy equipartition law applied to a collection of $3r$ independent harmonic oscillators.

Concept

In a classical harmonic solid, the high-temperature specific heat is a constant independent of temperature.

Quick Questions

Q18. Discuss why a quantum approach is needed to explain the low temperature limit of the heat capacity.

Answer. As shown in Fig. 3.10, the heat capacity should approach zero in the low temperature limit. This means that at $T=0$ K, strictly speaking, an infinitesimal variation of the temperature dT does not produce a change of the system internal energy ($dU = 0$).

From a classical point of view, the energy of a harmonic oscillator can continuously be varied from 0 to $+\infty$, thus any change in the temperature must lead to a variation of internal energy, implying $c_V \neq 0$. On the other hand, the quantum description relies on the quantized nature of the excitations of harmonic oscillators. If $\hbar\omega_0$ is the energy quantum for one of such oscillators, no internal energy variation may occur if $k_bT \ll \hbar\omega_0$. Indeed, in this case the thermal energy is not sufficient for the oscillator being excited, thus freezing it in its ground state.

The quantum calculation of the heat capacity of an insulator is now going to be performed within different approximations to (3.139).

Examples

The first attempt to calculate the quantum vibrational energy of an insulator is due to Einstein. In this model, the solid material is treated as a set of N harmonic three-dimensional oscillators, sharing the same frequency ω_0 . Therefore, such approximation relies on neglecting both the dependence of the frequencies on k , that is their dispersion, and on the band index: $\omega_s(\mathbf{k}) \simeq \omega_0$. As derived from (3.139), the energy becomes

$$U = \frac{3N\hbar\omega_0}{e^{\hbar\omega_0/(k_bT)} - 1}, \quad (3.142)$$

and the specific heat per particle is given by

$$c_V = \frac{1}{N} \left(\frac{\partial U}{\partial T} \right). \quad (3.143)$$

In the high-temperature limit, c_V reproduces the Dulong and Petit law. In the opposite $k_bT \ll \hbar\omega_0$ limit instead, it exponentially vanishes as $c_V \simeq 3\hbar\omega_0 e^{-\hbar\omega_0/(k_bT)}$.

From the example above, it is clear that a better approximation is required to correctly predict the experimental outcomes as $T \rightarrow 0$.

Examples

In the limit of very low temperatures, it is expected that only the low frequency excitations can contribute to the specific heat. In other words, the contribution of all the vibrational modes with $k_bT \ll \hbar\omega_s(\mathbf{k})$ can be ignored, because the least energy required for this modes to be excited is larger than thermal energy. This model is named after Debye. Within this approximation, the sum in (3.139) can be restricted only to the acoustic band $\omega_{ac}(\mathbf{k})$. In the $\mathbf{k} \rightarrow 0$ limit $\omega_{ac}(\mathbf{k}) \propto k$. Provided that $k_bT \gg \hbar c_s k$, the vibrational part of (3.139) can be cast in the form

$$U_{ph} = \int_0^{\omega_D} \frac{g_{ac}(\omega)\hbar\omega}{e^{\frac{\hbar\omega}{k_bT}} - 1} d\omega$$

$$= \frac{V}{(2\pi)^3} 4\pi\hbar \left(\frac{2}{c_t^3} + \frac{1}{c_r^3} \right) \int_0^{\omega_D} \frac{\omega^3}{e^{\frac{\hbar\omega}{k_b T}} - 1} d\omega. \quad (3.144)$$

Here, $g_{ac}(\omega)$ is the density of acoustical modes and the frequency ω_D is introduced so that

$$\int_0^{\omega_D} g_{ac}(\omega) d\omega = 3Nr. \quad (3.145)$$

This latter condition is required to ensure that $3Nr$ modes are contributing to the specific heat, otherwise the high temperature limit would be break up. It is easily shown that

$$\omega_D = \left[\frac{18\pi^2 Nr}{V} \frac{1}{\left(\frac{2}{c_t^3} + \frac{1}{c_r^3} \right)} \right]^{1/3}.$$

Setting $x \equiv \hbar\omega/(kT)$, (3.144) leads to

$$U_{ph} = \frac{V}{2\pi^2} \left(\frac{2}{c_t^3} + \frac{1}{c_r^3} \right) \left(\frac{k_b T}{\hbar} \right)^4 \int_0^{\frac{\Theta_D}{T}} \frac{x dx}{e^x - 1}, \quad (3.146)$$

with $\Theta_D \equiv \hbar\omega_D/k_b$. In the temperature range $\Theta_D/T \gg 1$ under consideration, the integral can be safely approximated by performing the $\Theta_D/T \rightarrow \infty$ limit. Thus, the specific heat per particle results to be

$$c_V = \frac{12\pi^4}{5} k_b \left(\frac{T}{\Theta_D} \right)^3. \quad (3.147)$$

The above derivation represents the Debye model for specific heat in solids.

Concept

In a quantum harmonic solid, the low-temperature specific heat vanishes according to a T^3 law on the scale of the Debye temperature Θ_D , embodying average information on the vibrational structure characteristic of the given crystal.

Table 3.2 lists and compares the Debye temperatures of selected cubic materials, as calculated from measurements of the average speeds of sound along the different crystallographic directions. The latter are inferred from elasticity data and from low-temperature thermal measurements.

In facts, measurements of the specific heat reflect the presence of all possible types of excitations, including for example electronic ones. Thus, the present analysis is suited to insulators. For metals instead, the electronic contribution has to be added, as it is derived in Chap. 4.

Table 3.2 Debye temperature Θ_D for the selected cubic structures in first column: comparison between values obtained from elastic data (third column) and those from thermal data (fourth column). The lattice parameters are also reported (second column) [13]

| Crystal | Lattice parameter (\AA) | Θ_D (from elastic data) | Θ_D (from thermal data) |
|---------|---------------------------------------|-----------------------------------|-----------------------------------|
| Na | 4.29 | 164 | 158 |
| Cu | 3.61 | 332 | 343 |
| Al | 4.05 | 403 | 428 |
| Si | 5.43 | 576 | 640 |
| Ge | 5.66 | 323 | 370 |
| LiF | 4.02 | 610 | 732 |
| NaCl | 5.63 | 289 | 321 |

Quick Questions

Q19. Consider the Debye model as applied to a 2D system. Discuss the power-law temperature behavior of specific heat at low temperatures $T \ll \Theta_D$.
Answer. Repeating the calculations performed for the 3D system and taking into account the different measure of integration, one finds $c_V \propto T^2$, yet on the scale of a 2D Debye temperature.

3.10 Thermodynamics of vibrational states: anharmonic effects

The present section is devoted to resolve the failure of the harmonic approximation, in accounting for finite values of thermodynamic response coefficients such as thermal expansion, conductivity and specific heat, which in turn describe the way a material responds to thermal driving. To this aim, the next-to-second order terms are considered in the expansion of the vibrational energy, representing beyond-harmonic or else anharmonic effects. The results obtained in the present section are in general valid for insulators for the reasons outlined at the end of the previous section, since for metals electronic effects are to be considered in addition to vibrational ones. The complete discussion is therefore referred to Chap. 4.

3.10.1 Corrections to the specific heat in classical limit

Anharmonicity introduces corrections to the vibrational contribution to specific heat. In facts,

Concept

The presence of significant anharmonic effects marks the behavior or even the existence of thermal properties of materials.

This can be seen in the simplest one-dimensional model, where anharmonic terms are included and worked out in the classical limit $T \gg \Theta_D$. Consider indeed the 1D model with up to fourth-order anharmonic interactions:

$$H = \frac{p^2}{2M} + U(x),$$

$$U(x) = ax^2 - bx^3 - cx^4. \quad (3.148)$$

The average energy \bar{E} is

$$\bar{E} = \frac{\int \int dx dp H e^{-\beta H}}{\int \int dx dp e^{-\beta H}}. \quad (3.149)$$

The term $e^{-\beta H}$ is calculated by expanding the anharmonic terms to lowest order. It is found

$$\bar{E} = k_b T + \left[\frac{3}{4} \frac{c}{a^2} + \frac{15}{16} \frac{b^2}{a^3} \right] (k_b T)^2, \quad (3.150)$$

depending on the coefficients of the anharmonic expansion (3.148). Calculation of the temperature derivative (3.150) to get the specific heat, leads to the conclusion that anharmonic terms determine a correction to the result of Dulong and Petit that is linear in the temperature.

Examples

In the following, the procedure set up for the determination of system response to external thermal driving is applied to the calculation of thermal expansion and of conductivity coefficients in anharmonic regime.

3.10.2 Thermal expansion

Thermal expansion coefficient α represents the system response to small temperature changes which drive a change in the system size. The coefficient α can be defined for small temperature variations, observing that while pressure is kept constant, a volume change is proportional to the volume itself and to the temperature change, that is $\Delta V = \alpha V \Delta T$. Using partial derivatives, this definition can be represented as:

$$\alpha = \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_P.$$

Materials more often expand in response to a temperature increase, so that typically $\alpha > 0$. Negative or null thermal expansion behavior are however possible and much wanted for structural, optical and electronic applications. Materials with negative thermal expansion shrink after an increase of temperature, at least in a given temperature range. Examples are some complex metal oxides, polymers, quartz and a number of zeolites. On the other hand, a Nickel-Iron alloy composed of 36% Nickel and 64% Iron has $\alpha \simeq 1.2 \cdot 10^{-6} \text{ K}^{-1}$ in the temperature range between 20° and 100° C : for its volume invariability to changes in such a wide temperature range it is named Invar, and it has provided its inventor Charles Guillaume with the Nobel prize.

The average increase in size $\langle x \rangle$ has been seen to vanish in harmonic approximation. In the 1D simplest model (3.148) with inclusion of anharmonic terms, one has instead:

$$\begin{aligned} \langle x \rangle &= \frac{\int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dx x e^{-\beta H}}{\int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dx e^{-\beta H}} \\ &\simeq \frac{\int_{-\infty}^{\infty} dx (x + \beta b x^4) e^{-\beta a x^2}}{\int_{-\infty}^{\infty} dx e^{-\beta a x^2}} \simeq \frac{3b}{4a^2} k_b T. \end{aligned} \quad (3.151)$$

The first term in second line in (3.151) is calculated in the same approximation as in (3.150). The expression of $\langle x \rangle$ is seen to depend on the coefficient b and on temperature T in much the same way as the specific heat does in the classical limit. A similar conclusion can be reached if the calculation were performed in the low-temperature quantum regime.

3.10.3 Thermal conductivity

Thermal conductivity χ_c represents the system response to small temperature gradients, which drive an energy current. To linear order in the temperature gradient, χ_c is defined by the relation

$$j(x) = \chi_c \left(-\frac{\partial T}{\partial x} \right),$$

the minus sign accounting for the fact that energy fluxes proceed from regions with larger to regions with lower temperatures. Phenomenological observations of thermal conductivity are often performed after placing a layer of selected materials between regions at different temperatures, thereby driving the heat flux to be measured. Heat transport occurs due to interaction mechanisms involving two different effective carriers: electrons and phonons. The interactions energies in turn depend on the location and thus on the temperature of the carrier. In the former case, electrons scatter across the layer in the two opposite directions, so that the average energy they carry on depends on the location where they have experimented the last interaction with the lattice. In the latter case, phonons interact among them due to the

anharmonic terms, they no longer being independent excitations as in the harmonic approximation.

In the simplest phenomenological model, phonons can be considered as quasi-particles moving with average velocity v and interacting with each other, the interactions being effectively embodied in the average time τ lasting between two subsequent phonon-phonon scattering events: the larger τ , the weaker the interactions and thus the anharmonic effects. The energy current $j(x) = -vU(x)$ written in terms of the energy density $U(x)$, can be expanded to linear order in the displacement x as $j(x) = -v\ell(\partial U/\partial x)$, $\ell \approx v\tau$ being the mean distance a vibrational excitation travels during time τ . The change of energy $U(x)$ is actually due to the presence of a temperature gradient, that is $(\partial U/\partial x) = (\partial U/\partial T(x))|_{x=\ell}(\partial T(x)/\partial x)$. Collecting all the relations together and considering that averaging the quadratic velocity in a isotropic 3D system introduces a factor $1/3$, the thermal conductivity can be expressed as:

$$\chi_c = \frac{v^2}{3} \tau n c_v. \quad (3.152)$$

This topic is reconsidered in Chap. 4, where the electronic contribution to thermal conductivity is included.

Quick Questions

Q20. Consider a material where anharmonic effects are vanishingly small.

Would such a material expand after being subjected to a temperature gradient?

Answer. No. The volume change of the material within the harmonic approximation has been found to be zero.

Q21. Consider the anharmonic potential (3.148), where a is assumed to be a positive quantity. How many critical points this anharmonic potential yields?

Answer. The condition $dU(x)/dx = 0$ yields to the equation $2ax - 3bx^2 - 4cx^3 = 0$, with solutions $x_1 = 0$ and $x_{2,3} = \left(-3b \pm \sqrt{9b^2 + 32ac}\right)/8c$. If $9b^2 + 32ac > 0$ the potential has three critical points. If $9b^2 + 32ac < 0$, or else if $c < -\frac{9b^2}{32a}$, only one critical point exists at $x = 0$, corresponding to a minimum. If c were vanishing, the solution with $x_1 = 0$ could not correspond to a minimum.

3.11 Electron-phonon interaction

In Chap. 2 and in the present Chapter, electronic and vibrational states have been introduced in the Born-Oppenheimer approximation, freezing any interaction between the two types of excitations. Electron-phonon interactions however manifest everywhere in solid state physics, so that it is now important to discuss this topic. The general form of the electron-phonon interaction is first introduced. Then, it is

discussed how effective attractive electron-electron interactions may come about in metals, that are originated by an electron-phonon interaction. This result represents the microscopic ingredient that explains the success of BCS theory for metallic superconductors. Additional important applications of electron-phonon interactions are treated in Chap. 5, where it is seen that phonons may assist and influence optical transitions and therefore optical properties of materials.

3.11.1 General form of the electron-phonon interaction

The general form of the electron-phonon interaction requires to go beyond the Born-Oppenheimer approximation. A way to introduce the interaction between electrons and phonons is to consider the whole Hamiltonian that includes electrons and ions and then proceed in a systematic manner. The starting Hamiltonian is

$$\begin{aligned}
 H = & -\sum_{\alpha} \frac{\hbar^2}{2M_{\alpha}} \nabla_{\alpha}^2 - \sum_i \frac{\hbar^2}{2m} \nabla_i^2 \\
 & + \sum_{\alpha, i} V_{\alpha}(\mathbf{R}_{\alpha} - \mathbf{r}_i) + \sum_{i < j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \\
 & + \sum_{\alpha < \beta} U_{\alpha, \beta}(\mathbf{R}_{\alpha} - \mathbf{R}_{\beta}). \tag{3.153}
 \end{aligned}$$

Here, \mathbf{R}_{α} and \mathbf{r}_i are the positions of ion α and electron i , respectively, $V_{\alpha}(\mathbf{R}_{\alpha} - \mathbf{r}_i)$ represents the pair interaction potential between them, and $U_{\alpha, \beta}(\mathbf{R}_{\alpha} - \mathbf{R}_{\beta})$ the interaction between two ions α and β . In (3.153), the contribution of core electrons with other electrons and ions is assumed to have been already embodied in V_{α} and $U_{\alpha, \beta}$. This is an approximation in fact, though reasonable enough: as it has been seen while developing the band theory, the core electron orbitals are not significantly modified when the atom is placed in the crystal. In any event, their contribution to V_{α} and $U_{\alpha, \beta}$ can be traced back to preliminar calculations at the level of atomic physics. As a consequence, in (3.153) the nuclei vibrations, implicitly given by the time evolution of \mathbf{R}_{α} , occur with the electrons that rigidly follow them. This is not necessarily the case of course: if it is not, this effect is to be embodied from the start in V_{α} and $U_{\alpha, \beta}$.

The first step to determine the eigenfunction spectrum of (3.153) is the Born-Oppenheimer approximation, where the electronic problem is solved while the ions are considered fixed in their equilibrium positions and the eigenvalues $E_n(\mathbf{R}_{\alpha}, \mathbf{R}_{\beta}, \dots)$, that contain also the term $\sum_{\alpha < \beta} U_{\alpha, \beta}(\mathbf{R}_{\alpha} - \mathbf{R}_{\beta})$, depend on the ionic coordinates. Then, the interaction potential between the ions is taken to be $E_n(\mathbf{R}_{\alpha}, \mathbf{R}_{\beta}, \dots)$. The problem of lattice vibrations is solved by expanding $E_n(\mathbf{R}_{\alpha}, \mathbf{R}_{\beta}, \dots)$ up to second order in the displacements. The equilibrium positions are the minima of $E_n(\mathbf{R}_{\alpha}, \mathbf{R}_{\beta}, \dots)$, that are found by means of

$$\nabla_{\mathbf{R}_{\alpha}} E_n(\mathbf{R}_{\alpha}, \mathbf{R}_{\beta}, \dots) = 0, \forall \alpha.$$

Therefore, the energy spectrum of electrons is found after placing the ions at the equilibrium positions \mathbf{R}_α^0 . The vibrational spectrum is instead found after solving the Schrödinger of the ions in the harmonic potential derived from the expansion of $E_n(\mathbf{R}_\alpha, \mathbf{R}_\beta, \dots)$ to second order in the displacement around the equilibrium positions \mathbf{R}_α^0 . The remaining terms in (3.153) are

$$\begin{aligned} V_\alpha(\mathbf{R}_\alpha - \mathbf{r}_i) - V_\alpha(\mathbf{R}_\alpha^0 - \mathbf{r}_i) &= V_\alpha(\mathbf{R}_\alpha^0 + \mathbf{s}_\alpha - \mathbf{r}_i) - V_\alpha(\mathbf{R}_\alpha^0 - \mathbf{r}_i) \\ &= -\nabla_{\mathbf{r}_i} V_\alpha(\mathbf{R}_\alpha^0 - \mathbf{r}_i) \cdot \mathbf{s}_\alpha + \frac{1}{2} \nabla_{\mathbf{r}_i} \nabla_{\mathbf{r}_j} V_\alpha(\mathbf{R}_\alpha^0 - \mathbf{r}_i) \cdot \mathbf{s}_\beta \mathbf{s}_\gamma + \dots \\ &= H_{e-ph}(\mathbf{r}_i, \mathbf{s}_\alpha). \end{aligned} \quad (3.154)$$

Here $\nabla_{\mathbf{r}_i} V_\alpha$ and $\nabla_{\mathbf{r}_i} \nabla_{\mathbf{r}_j} V_\alpha$ represent a vector and a tensor respectively, built up from first and second order gradients of V_α with respect to \mathbf{r}_i . $H_{e-ph}(\mathbf{r}_i, \mathbf{s}_\alpha)$ is thus structured as a sum of products of \mathbf{r}_i -dependent functions that are multiplied times powers of the displacements to different orders: the lowest is a linear order. In addition, the remaining terms, $E_n^{\text{res}}(\mathbf{R}_\alpha, \mathbf{R}_\beta, \dots)$ in the expansion of $E_n(\mathbf{R}_\alpha, \mathbf{R}_\beta, \dots)$ yield

$$E_n^{\text{res}}(\mathbf{R}_\alpha, \mathbf{R}_\beta, \dots) = \varepsilon_n(\mathbf{R}_\alpha^0, \mathbf{R}_\beta^0, \dots) + \bar{H}_{\alpha,\beta,\dots}(\mathbf{s}_\alpha, \mathbf{s}_\beta, \dots). \quad (3.155)$$

Here, the lowest order term of $\bar{H}_{\alpha,\beta,\dots}(\mathbf{s}_\alpha, \mathbf{s}_\beta, \dots)$ is third order in the displacements.

In conclusion, the starting Hamiltonian (3.153) is substituted by

$$\begin{aligned} H &= -\sum_i \frac{\hbar^2}{2m} \nabla_i^2 + \sum_{\alpha,i} V_\alpha(\mathbf{R}_\alpha^0 - \mathbf{r}_i) + \sum_{i < j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \\ &+ \sum_{s,\mathbf{q}} \hbar \omega_s(\mathbf{q}) \left(b_{s,\mathbf{q}}^+ b_{s,\mathbf{q}} + \frac{1}{2} \right) + \sum_{i,\alpha,\beta,\dots} H_{e-ph}(\mathbf{r}_i, \mathbf{s}_\alpha, \mathbf{s}_\beta, \dots) \\ &+ \sum_{\alpha,\beta,\dots} \bar{H}_{\alpha,\beta,\dots}(\mathbf{s}_\alpha, \mathbf{s}_\beta, \dots) + E_n(\mathbf{R}_\alpha^0, \mathbf{R}_\beta^0, \dots), \end{aligned} \quad (3.156)$$

whose terms can be interpreted as follows:

Properties

P1.

$$H_{el} = -\sum_i \frac{\hbar^2}{2m} \nabla_i^2 + \sum_{\alpha,i} V_\alpha(\mathbf{R}_\alpha^0 - \mathbf{r}_i) + \sum_{i < j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|}$$

represents the electronic Hamiltonian with the ions at the equilibrium positions, as discussed in Chap. 2 .

P2.

$$H_{ph} = \sum_{s,\mathbf{q}} \hbar \omega_s(\mathbf{q}) \left(b_{s,\mathbf{q}}^\dagger b_{s,\mathbf{q}} + \frac{1}{2} \right)$$

represents the phonon Hamiltonian, as discussed in Chap. 3 .

P3.

The quantity $\sum_{i,\alpha,\beta,\dots} H_{e-ph}(\mathbf{r}_i, \mathbf{s}_\alpha, \mathbf{s}_\beta, \dots)$ involves both the vibrational and electronic variables: it represents the electron-phonon interaction .

P4.

$\sum_{\alpha,\beta,\dots} \bar{H}_{\alpha,\beta,\dots}(\mathbf{s}_\alpha, \mathbf{s}_\beta, \dots)$ represents the anharmonic terms .

P5.

$E_n(\mathbf{R}_\alpha^0, \mathbf{R}_\beta^0, \dots)$ represents the term originating the cohesive energy, as discussed in Chap. 1 .

Equation (3.156) therefore yields the whole Hamiltonian of the system composed of ions and electrons, whose electron-phonon interaction term has been identified.

As already done for the Hamiltonian term of independent phonons

$$H_{ph} = \sum_{s,\mathbf{q}} \hbar \omega_s(\mathbf{q}) \left(b_{s,\mathbf{q}}^\dagger b_{s,\mathbf{q}} + \frac{1}{2} \right),$$

the electron-phonon term can as well be cast in terms of the phonon quantum operators $b_{s,\mathbf{q}}^\dagger$ and $b_{s,\mathbf{q}}$. As shown in Appendix 3.19, one has

$$H_{e-ph} = \sum_{i,\mathbf{q},s} \sqrt{\frac{\hbar}{2\omega_s(\mathbf{q})}} (b_{s,-\mathbf{q}}^\dagger + b_{s,\mathbf{q}}) F(\mathbf{q}, s, \mathbf{r}_i), \quad (3.157)$$

with

$$F(\mathbf{q}, s, \mathbf{r}_i) = \sum_\mu \sqrt{\frac{1}{M_\mu}} \mathbf{e}_\mu(\mathbf{q}, s) \cdot \sum_m \nabla_{\mathbf{r}_i} V_\mu(\mathbf{R}_{m,\mu} - \mathbf{r}_i) e^{i\mathbf{q} \cdot \mathbf{R}_m}. \quad (3.158)$$

In the next section this result is applied to show that the electron-phonon interaction leads to an effective electron-electron attractive interaction.

3.11.2 Attractive interaction between electrons in metals

Consider a gas of independent electrons, interacting only with phonons. The Hamiltonian $H = H_e + H_{ph} + H_{e-ph}$ is composed of the kinetic electron part $H_e =$

$\sum_i p_i^2/(2m)$, the free phonon term $H_{ph} = \sum_{\mathbf{q}} \hbar \omega(\mathbf{q}) b_{\mathbf{q}}^\dagger b_{\mathbf{q}}$, and the electron-phonon contribution H_{e-ph} (3.157). The electron term H_e can be treated on the same footing than the phonon one H_{ph} , that is in the creation and destruction operators representation instead than in Schrödinger representation. Using the tools developed in Appendix 2.14, one finds that $H_e = \sum_{\mathbf{k}} \varepsilon(\mathbf{k}) c_{\mathbf{k}}^\dagger c_{\mathbf{k}}$, where $c_{\mathbf{k}}$ ($c_{\mathbf{k}}^\dagger$) is the fermion operator destroying (creating) a particle in state with momentum \mathbf{k} and kinetic energy $\varepsilon(\mathbf{k}) = \hbar^2 k^2/(2m)$.

The treatment of the electron-phonon interaction requires some additional simplification. To this aim, notice that the quantity $F(\mathbf{q}, s, \mathbf{r}_i)$ in (3.158) can be recast in the form $F(\mathbf{q}, s, \mathbf{r}_i) = e^{i\mathbf{q} \cdot \mathbf{r}_i} D_0(\mathbf{q}, s)$. Thus, H_{e-ph} can be expressed as

$$H_{e-ph} = \sum_{i,s,\mathbf{q}} e^{i\mathbf{q} \cdot \mathbf{r}_i} D(\mathbf{q}, s) \left(b_{-\mathbf{q},s}^\dagger + b_{\mathbf{q},s} \right). \quad (3.159)$$

Consider now the case in which only one acoustic phonon mode is considered, so that (3.159) becomes $H_{e-ph} = \sum_{i,\mathbf{q}} e^{i\mathbf{q} \cdot \mathbf{r}_i} D(\mathbf{q}) \left(b_{-\mathbf{q}}^\dagger + b_{\mathbf{q}} \right)$. Finally, the term $e^{i\mathbf{q} \cdot \mathbf{r}_i}$ has to be expressed too in terms of fermion creation and destruction operators, leading to the operator form $e^{i\mathbf{q} \cdot \mathbf{r}_i} = \sum_{\mathbf{k}} c_{\mathbf{k}+\mathbf{q}}^\dagger c_{\mathbf{k}}$. In essence, when applied to the ground state of the many electron system, this operator creates one electron with wavevector $\mathbf{k} + \mathbf{q}$ after that one electron has been destroyed with wavevector \mathbf{k} : that is, it creates a fluctuation in the density of electrons characterized by wavevector \mathbf{q} .

As a result, the Hamiltonian of the system of electrons non-interacting among themselves and interacting with one acoustical phonon, is

$$H = \sum_{\mathbf{k}} \varepsilon(\mathbf{k}) c_{\mathbf{k}}^\dagger c_{\mathbf{k}} + \sum_{\mathbf{q}} \hbar \omega(\mathbf{q}) b_{\mathbf{q}}^\dagger b_{\mathbf{q}} + \sum_{\mathbf{k},\mathbf{q}} D(\mathbf{q}) c_{\mathbf{k}+\mathbf{q}}^\dagger c_{\mathbf{k}} \left(b_{-\mathbf{q}}^\dagger + b_{\mathbf{q}} \right). \quad (3.160)$$

Eigenvalues and eigenfunctions are now to be found for (3.160). To this aim, the following procedure is used to eliminate the boson operators and obtain an effective all-electron Hamiltonian. One first notices that $H = H_0 + H'$ can be decomposed into non-interacting $H_0 = H_e + H_{ph}$ and interacting $H' \equiv H_{e-ph}$ parts. Then, an unitary transformation S is applied to H , such that $\tilde{H} = e^{-S} H e^S$ just contains quadratic $D(\mathbf{q})$ terms. Now, one has that

$$\begin{aligned} \tilde{H} &= H + [H, S] + \frac{1}{2} [[H, S], S] + \dots \\ &= H_0 + (H' + [H_0, S]) + \frac{1}{2} [(H' + [H_0, S]), S] + \frac{1}{2} [H', S] \dots \end{aligned} \quad (3.161)$$

Thus, S must be linear in $D(\mathbf{q})$ and the equation

$$H' + [H_0, S] = 0 \quad (3.162)$$

must be satisfied. To find out the expression for S , one tries a form similar to that of the electron-phonon Hamiltonian term, that is:

$$S = \sum_{\mathbf{k}, \mathbf{q}} D(\mathbf{q}) \left(\alpha_{\mathbf{k}, \mathbf{q}} b_{-\mathbf{q}}^\dagger + \beta_{\mathbf{k}, \mathbf{q}} b_{\mathbf{q}} \right) c_{\mathbf{k}+\mathbf{q}}^\dagger c_{\mathbf{k}}, \quad (3.163)$$

and then identifies the unknown functions $\alpha_{\mathbf{k}}$ and $\beta_{\mathbf{k}}$ by applying the condition (3.162). This yields to the result:

$$\begin{aligned} \alpha_{\mathbf{k}, \mathbf{q}}^{-1} &= \varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k} + \mathbf{q}) - \hbar\omega(\mathbf{q}) \\ \beta_{\mathbf{k}, \mathbf{q}}^{-1} &= \varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k} + \mathbf{q}) + \hbar\omega(\mathbf{q}). \end{aligned} \quad (3.164)$$

\tilde{H} is then given to the lowest order as $\tilde{H} = H_0 + [H', S]/2$. Collecting all the above results together, one may finally eliminate the phonon operators and find out the result:

$$\tilde{H} = \sum_{\mathbf{k}} \varepsilon(\mathbf{k}) c_{\mathbf{k}}^\dagger c_{\mathbf{k}} + \sum_{\mathbf{q}, \mathbf{k}, \mathbf{k}'} \frac{D^2(\mathbf{q}) \hbar\omega(\mathbf{q})}{[\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k} - \mathbf{q})]^2 - \hbar^2 \omega^2(\mathbf{q})} c_{\mathbf{k}'+\mathbf{q}}^\dagger c_{\mathbf{k}-\mathbf{q}}^\dagger c_{\mathbf{k}}. \quad (3.165)$$

Notice that now \tilde{H} actually represents the average $\langle 0 | \tilde{H} | 0 \rangle$ over the ground state $|0\rangle$ of the phonon field. In essence, the original system composed of electrons and phonons interacting only among them has been mapped into an electron system with an effective interaction: its size is dictated by the electron-phonon coupling strength D^2 , and its shape is determined by the electron and the phonon free-particle energies $\varepsilon(\mathbf{k})$ and $\hbar\omega_{\mathbf{k}}$. In fact, the second term in (3.165) describes a situation in which two electron-density fluctuations interact by exchanging one phonon quantum with wavevector \mathbf{q} .

Here is therefore the main result of this section. In the presence of a (acoustical) phonon, originally independent electrons are effectively seen to interact among them via the term $D^2 \hbar\omega(\mathbf{q}) / \left[[\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k} - \mathbf{q})]^2 - \hbar^2 \omega^2(\mathbf{q}) \right]^{-1}$: this term becomes negative whenever $|\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k} - \mathbf{q})| < \hbar\omega(\mathbf{q})$, thereby representing an attractive interaction. The attractive electron-electron interaction is entirely due to quantum effects and has no classical analogue. For electrons close to the Fermi level and using the Debye model, the attractive term acts when their energy ε is in the range $\varepsilon_f - \hbar\omega_D < \varepsilon < \varepsilon_f + \hbar\omega_D$ with ω_D Debye frequency. Superconductivity in metallic compounds is a sharp manifestation of this specific coupling.

Notice also that the above procedure is quite general and similar results can be obtained if the boson field mediating the pairing is substituted with other boson excitations. These might for example be optical phonons, magnons, and the like: the form of the Hamiltonian (3.160) would clearly be preserved.

3.12 Experimental methods to determine vibrational spectra

Experimental methods and techniques suited to determine the structure of macroscopic and nanoscopic crystals have been discussed in Chap. 1. X-ray elastic scat-

tering has been discussed, where the energy and frequency of scattered radiation are essentially the same as the those of the incoming waves. Elastic scattering is suited to determine geometric structures of crystals and nanostructures, where one is in essence interested in the arrangement of atoms at their equilibrium positions. If the excitations of the crystal structure were to be considered, the concept had been anticipated according to which inelastic scattering is to be used. In inelastic scattering indeed, the change in energy is precisely provided by the energy of the probed excitation. Experimental methods to determine phonon excitations in crystals need indeed to use probes that couple to the density of ions, being able to determine density fluctuations, namely a sound wave.

To this aim, the simplest idea is to use mechanical waves as probes. Also, beams of heavy particles can be used such as thermal neutrons or else even hard, that is high-energy, X-rays that inelastically scatter against nuclei. As a thumb rule, either particle-like or wave-like probes should be selected to possess wavelengths and energies comparable to the characteristic size and energy scales of the system under study. Though the theories for neutron and X-ray inelastic scattering work in similar ways, some additional care is required when using X radiation: first, it interacts with electrons besides nuclei; second, photons with very high energies in the range 10-20 keV are needed to explore the whole Brillouin zone, and the needed resolutions in the 0.1-1 meV range would correspond to challenging accuracies of the order of $10^{-7} - 10^{-8}$. These are more recently made available from very bright and focused sources of synchrotron radiation, that is produced in large facilities by electrons rapidly bending in the magnetic field of a synchrotron ring. Finally, vibrational modes can be excited also after absorption of infrared light, that is with wavelength suited to the meV phonon energies, or else after inelastic scattering of light that is named from Raman. Therefore,

Concept

Determination of phonon spectra can be obtained by means of mechanical-wave probes, or X-ray inelastic scattering though from bright synchrotron sources, or neutron inelastic scattering, or else by light infrared absorption or Raman inelastic scattering processes.

Details on the specific mechanisms are postponed to Chap. 5, where the theoretical tools are built up that can deepen the understanding of the working principles. Here just a sketch of the different possibilities is illustrated. The interested student is also referred to [15, 16].

3.12.1 Sound propagation

Elastic constants of crystals can be determined after measuring the speed of sound propagation for longitudinal and transverse settings. Mechanical, sound waves can

be originated by applying to the crystal surface a piezoelectric transducer, transforming electric into mechanical signals. The generated waves, typically ultrasound waves with frequencies of the order of 10 MHz, propagate into the crystal, are echoed back from the opposite crystal edge, and are detected by the piezoelectric transducer working in reverse setting.

3.12.2 Inelastic Neutron Scattering

A very common probe of vibrational states in crystals and nanostructures is inelastic scattering of thermal neutrons, that are neutral particles with mass and energies $k_b T$, corresponding to de Broglie wavelengths comparable to typical distances between atoms inside the crystal. The use of neutral particles is suited to probe local, short-range motion of atoms nuclei. In practice, consider a neutron capable of interacting with the atoms nuclei vibrating around their equilibrium lattice positions. The Hamiltonian which describes the states of the system is

$$H = \frac{p^2}{2M_n} + \sum_{m,\mu} V_\mu [\mathbf{r} - \mathbf{R}_m - \mathbf{d}_\mu - \mathbf{s}_{m,\mu}(t)]. \quad (3.166)$$

Here, $\mathbf{s}_{m,\mu}(t)$ is as usual the displacement of atomic species μ around equilibrium position \mathbf{d}_μ within the cell \mathbf{R}_m . In expression (3.166), the terms describing the motion of ions around their equilibrium positions is assumed to be already determined and embodied in the knowledge of the time-dependent behavior of each displacement $\mathbf{s}_{m,\mu}$. A neutron interacts solely with the atoms nuclei, thus the functions $V_\mu [\mathbf{r} - \mathbf{R}_m - \mathbf{d}_\mu - \mathbf{s}_{m,\mu}(t)]$ are tightly localized around them. In the limit of small displacements $|\mathbf{s}_{m,\mu}(t)|$ on the scale of average distances between nuclei in the lattice, (3.166) can be expanded to linear order in $\mathbf{s}_{m,\mu}$. Expansion of $V_\mu [\mathbf{r} - \mathbf{R}_m - \mathbf{d}_\mu - \mathbf{s}_{m,\mu}(t)]$ to first order in the displacements yields the simplified form:

$$H = \frac{p^2}{2M_n} + \sum_{m,\mu} [V_\mu(\mathbf{r} - \mathbf{R}_m - \mathbf{d}_\mu) - \nabla_{\mathbf{r}} V_\mu(\mathbf{r} - \mathbf{R}_m - \mathbf{d}_\mu) \cdot \mathbf{s}_{m,\mu}(t)]. \quad (3.167)$$

The structure of Hamiltonian (3.167) is useful, since the first two terms have the form of a time-independent Hamiltonian composed of kinetic and potential pieces. The third term embodies instead the time dependence through the displacements $\mathbf{s}_{m,\mu}(t)$.

In the subsequent calculation, the first term acts as an unperturbed Hamiltonian $H_0 = p^2/(2M_n)$ and the remaining two terms as perturbations. In appendix 3.18 the probability per unit time is calculated, with which an unperturbed initial state ψ_{in} performs a transition to a final excited state ψ_f of H_0 , as a consequence of an external time-dependent perturbation $\bar{H}(t)$. The calculation is performed by considering the Hamiltonian H_0 of systems with one or more particles, with the operator $\bar{H}(t)$

acting in an arbitrary time interval $(0, \tau)$. Expression (3.225) provides the transition probability per unit time to first perturbative order when the matrix elements of the perturbation $H'_{ij}(t)$ are of the form $H'_{ij}(t) = W_{ij}(e^{i\omega t} + e^{-i\omega t})$. The time-dependent vibrational displacements are assumed to be known, and are generally provided in the classical-mechanics form:

$$\mathbf{s}_{m,\mu} = \sum_{\mathbf{k},s} \left[b_{\mathbf{k},s}^* e^{i\omega_s(\mathbf{k})t} + b_{\mathbf{k},s} e^{-i\omega_s(\mathbf{k})t} \right] \sqrt{\frac{M}{M_\mu}} \mathbf{e}_{s,\mu}(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{R}_m}, \quad (3.168)$$

with $b_{\mathbf{k},s}(t)$ and $b_{\mathbf{k},s}^*$ constants to be determined from the initial conditions. It is also characterized by validity conditions similar to those commonly used in the study of electromagnetic transitions between electronic states in a quantum system within a quasi-classical approximation for the radiation, that is with the radiation treated in terms of Maxwell equations.

The initial and final unperturbed states can be represented as $\psi_i = V^{-\frac{1}{2}} e^{i\mathbf{k}_i \cdot \mathbf{r}}$ and $\psi_f = V^{-\frac{1}{2}} e^{i\mathbf{k}_f \cdot \mathbf{r}}$, with energies $E_i = \hbar^2 k_i^2 / (2M)$ and $E_f = \hbar^2 k_f^2 / (2M)$. The time-independent perturbation term in (3.167) leads to the probability per unit time for the transition from $\psi_i \rightarrow \psi_f$:

$$\begin{aligned} P &= \frac{2\pi}{\hbar} \langle \psi_i | \sum_{m,\mu} V_\mu(\mathbf{r} - \mathbf{R}_m - \mathbf{d}_\mu) | \psi_f \rangle \delta(E_i - E_f) \\ &= \frac{2\pi}{\hbar} \sum_{m,\mu} \tilde{V}_\mu(\Delta \mathbf{k}) e^{i\Delta \mathbf{k} \cdot (\mathbf{R}_m - \mathbf{d}_\mu)} \delta(E_i - E_f). \end{aligned} \quad (3.169)$$

Here, $\tilde{V}_\mu(\Delta \mathbf{k})$ is the spatial Fourier transform of $V_\mu(\mathbf{r})$ and $\Delta \mathbf{k} \equiv \mathbf{k}_f - \mathbf{k}_i$. In the case of elastic scattering, one finds that $\sum_m e^{i\Delta \mathbf{k} \cdot \mathbf{R}_m}$ is significantly different from zero only if $\Delta \mathbf{k} = \mathbf{G}$, while the elastic condition $E_i = E_f$ corresponds to state that $k_i = k_f$. Besides, the term $\sum_\mu \tilde{V}_\mu(\Delta \mathbf{k}) e^{-i\Delta \mathbf{k} \cdot \mathbf{d}_\mu}$ appearing in (3.169) is similar to a geometrical structure factor. Collecting these remarks together, one may conclude that the perturbative term under consideration leads to the same results as in X-ray elastic scattering.

The term to first order in the displacements can be treated in similar manner. The following selection rules are found, in order to have non-vanishing matrix elements:

$$\begin{aligned} \mathbf{k}_f - \mathbf{k}_i &= \mathbf{q} + \mathbf{G} \\ \frac{\hbar^2 k_f^2}{2M_n} - \frac{\hbar^2 k_i^2}{2M_n} &= \pm \hbar \omega_s(\mathbf{q}) \end{aligned} \quad (3.170)$$

These expressions are especially relevant. In essence, given \mathbf{k}_i , \mathbf{k}_f , E_i and E_f , the wavevector \mathbf{q} and the energy $\hbar \omega_s(\mathbf{q})$ of the probed phonon are univocally determined from the first and second equation, respectively. Notice that all the perturbative terms have been analyzed separately and independently of each other. This strategy turns out to be reasonable a posteriori, since each of the perturbative terms shows up as a peak in the transition probability, located at a different energy.

To second order in perturbation theory, the following selection rules are found:

$$\mathbf{k}_f - \mathbf{k}_i = \mathbf{q} + \mathbf{q}' + \mathbf{G}$$

$$\frac{\hbar^2 k_f^2}{2M_n} - \frac{\hbar^2 k_i^2}{2M_n} = \pm \hbar \omega_s(\mathbf{q}) \pm \hbar \omega_s(\mathbf{q}'). \quad (3.171)$$

These rules do not fix \mathbf{q} and \mathbf{q}' in a univocal manner. Thus, the variations in energy that the neutron might acquire as a result of inelastic scattering, show up as a broad background. The same conclusion can be reached after analyzing the terms with higher order than the first. Thus,

Concept

Measurements of energy loss of incoming neutrons after they undergo inelastic scattering against vibrating nuclei, generally manifest as a background with better defined peaks on top of it due to first-order transitions.

This general structure of the inelastic neutron spectra is clearly visible in experiments, as displayed in Fig. 3.11. The finite width of the peaks reflects the non-stationary conditions of phonon excitations, they being indeed interacting among themselves due to anharmonic terms.

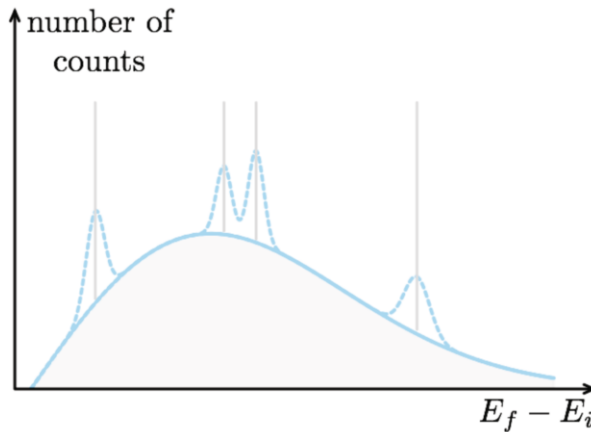


Fig. 3.11 Experimental determination of vibrational spectra by inelastic neutron scattering. Qualitative behavior of the transition probability of neutrons from an initial to a final state. The finite-width peaks are clearly visible, due to first-order transitions and anharmonic effects. The peaks are located at the different energies of different phonon modes, on top of a broad background due to higher-order transition terms

The same problem can be studied within a full quantum-mechanical formulation. The Hamiltonian to be considered is

$$H = \frac{p^2}{2M_n} + \sum_{\mathbf{q},s} \hbar \omega_s(\mathbf{q}) \left(b_{\mathbf{q},s}^\dagger b_{\mathbf{q},s} + \frac{1}{2} \right) + \sum_{m,\mu} [V_\mu(\mathbf{r} - \mathbf{R}_m - \mathbf{d}_\mu) - \nabla_{\mathbf{r}} V_\mu(\mathbf{r} - \mathbf{R}_m - \mathbf{d}_\mu) \cdot \mathbf{s}_{m,\mu}], \quad (3.172)$$

having the same structure as that in (3.167), though the physical meaning of the symbols is significantly different. First, Hamiltonian (3.172) represents the whole system of one neutron and the lattice vibrations. Second, $\mathbf{s}_{m,\mu}$ is now the quantum operator describing the atomic displacement. This can be expressed according to (3.113) along with the connection between the normal-mode variables and the creation and destruction operators:

$$\mathbf{s}_{m,\mu} = \sum_{\mathbf{q},s} \sqrt{\frac{2\hbar}{\omega_s(\mathbf{q})}} (b_{-\mathbf{q},s}^\dagger + b_{\mathbf{q},s}) \frac{\mathbf{e}_{s,\mu}(\mathbf{q})}{\sqrt{M_\mu}} e^{i\mathbf{q} \cdot \mathbf{R}_m}. \quad (3.173)$$

The initial and final states are here given by $\psi_i = V^{-\frac{1}{2}} e^{i\mathbf{k}_i \cdot \mathbf{r}} |..n_{\mathbf{q}s}...\rangle$ and $\psi_f = V^{-\frac{1}{2}} e^{i\mathbf{k}_f \cdot \mathbf{r}} |..n'_{\mathbf{q}s}...\rangle$, with energies

$$E_i = \hbar^2 k_i^2 / (2M_n) + \sum_{\mathbf{q}s} \hbar \omega_s(\mathbf{q}) n_{\mathbf{q}s}$$

and

$$E_f = \hbar^2 k_f^2 / (2M_n) + \sum_{\mathbf{q}s} \hbar \omega_s(\mathbf{q}) n'_{\mathbf{q}s}.$$

Here, $n_{\mathbf{q}s}$ and $n'_{\mathbf{q}s}$ are the occupation numbers of the initial and final phonon states. Application of the procedures for first and second-order perturbation theory yield once again the same selection rules, as in (3.169), (3.170) and (3.171). One additional information emerges from the present quantum treatment: as required by (3.169), the initial and final vibrational states are to coincide, that is $n_{\mathbf{q}s} = n'_{\mathbf{q}s}$, $\forall(\mathbf{q},s)$. Expression (3.170) instead requires that the vibrational and final states coincide, except for the possibility of allowing for a value of (\mathbf{q},s) for which $n'_{\mathbf{q}s} = n_{\mathbf{q}s} \pm 1$. In the case of second order transitions, the final vibrational state may differ from the initial one by two phonons.

3.12.3 Infrared Spectroscopy

Infrared Spectroscopy (IR) is based on the concept of absorption of light by a vibrating molecule or group of ions in a crystal cell, exciting the sample into a higher-energy vibrational state. The energy of the photon has to match the energy difference between the two vibrational states, that is typically in the infrared range. Lower

energy photons, in the microwave range, may instead excite rotational molecular states. For the vibrational-excitation event to occur and the sample be IR-active, a change in dipole moment μ is required after a change in the normal coordinate Q . The intensity of an infrared absorption peak scales as $I_{IR} \propto (\partial\mu/\partial Q)^2$. Therefore,

Concept

A molecule with same atoms cannot be IR-active, while molecules with different types of atoms are usually IR-active, even if the dipole moment is induced.

Carbon dioxide CO_2 is such an example: it is characterized by inversion symmetry but may have antisymmetric displacements, corresponding to an induced dipole moment.

Anharmonic effects in the vibrational potential, as described in Sec. 3.10, are essential to account for crucial characteristics of IR spectra. One example is the possibility of observing overtones, that is transitions between non-adjacent vibrational states with changes Δn in the quantum number n like $\Delta n = \pm 2, \dots$: these would be forbidden in a purely harmonic oscillator where $\Delta n = \pm 1$. Also, transitions to higher vibrational states are by far less likely and therefore less intense, the spacing between adjacent vibrational levels reducing with increasing energy up to dissociation.

Chap. 5 contains an extended treatment of the theoretical tools needed to understand absorption, scattering and reflection of light, and their relationship with the determination of the frequencies of longitudinal and transverse acoustical and optical phonons.

3.12.4 Raman Spectroscopy

Vibrational states can be probed as well by means of inelastic light-scattering instead than absorption processes of infrared-energy photons. In inelastic light-scattering processes, the state is excited into higher-energy, virtual states, where it immediately decays via emission of a photon. Light absorption processes involve one photon, light-scattering two of them. Photon energies employed to excite vibrational molecular states are in the infrared range for absorption processes. In inelastic scattering processes where part of the photon energy is indeed lost, the photon energies are instead higher than the typical vibrational energies, that is in the near-infrared or ultraviolet range.

The Rayleigh scattering illustrated in Chap. 2 is one example of light scattering, elastic though, where the emitted photon has the same energy as the incoming photon. If instead after the collision the energy of the emitted photon changes with respect to the incoming one, the process is inelastic. The energy difference between the inelastically scattered and the incident photon is precisely the energy difference between two vibrational states. This effect is named after the indian physicist Raman. Notice that when involving acoustical phonons, the same process is actually named after Brillouin. In particular, energy and momentum conservation laws imply that the energy $\hbar\omega$ and momentum \mathbf{k} of the incident photon, the energy $\hbar\omega'$ and

momentum \mathbf{k}' of the emitted photon, and the energy $\hbar\Omega(\mathbf{K})$ and momentum \mathbf{K} of the phonon be related by

$$\begin{aligned}\hbar\omega' &= \hbar\omega \pm \hbar\Omega(\mathbf{K}) \\ \hbar n\mathbf{k}' &= \hbar n\mathbf{k} \pm \hbar\mathbf{K},\end{aligned}\tag{3.174}$$

with n the sample index of refraction. The plus sign refers to phonon absorption and the minus to phonon emission. In essence,

Concept

A vibrational mode is Raman active if the structure polarizability changes during vibration.

That is, after a change in normal coordinate Q : the Raman signal intensity $I_R \propto I_0 \nu^4 N (\partial\alpha/\partial Q)^2$ depends on the change of polarizability during the vibration, besides the incident light intensity I_0 , the number N of scattering structures in the given vibrational state, and the fourth power of the frequency ν of the exciting laser [16].

Two different situations may thus occur in (inelastic) Raman scattering, as depicted in the top panel of Fig. 3.12. The vibrational energy of the molecule can be increased after the collision as in Fig. 3.12 (a): the energy of the scattered photons is decreased by the same amount and is detected at longer wavelengths, as represented by the cyan arrow. The corresponding structures in the Raman spectrum, as displayed in the bottom panel, are peaks shifted towards lower energies, namely the Stokes shift. The vibrational energy of the molecule can instead be decreased after the collision, as in Fig. 3.12 (c): the energy of the scattered photons is increased by the same amount and is detected at shorter wavelengths. The corresponding structures in the Raman spectrum, as displayed in the bottom panel of the figure, are peaks shifted towards higher energies, namely anti-Stokes shift. In order for this even to occur, phonons are to be already excited into higher vibrational states before the collision with the incident photon, as it may occur because of thermal effects according to Boltzmann statistics. Therefore, anti-Stokes events are less likely than Stokes ones, and the corresponding peaks have lower height. In fact, Stokes peaks involve the creation of one phonon and the peak intensity $I(\omega + \Omega) \propto \langle n_{\mathbf{K}} + 1 \rangle$ in terms of the average phonon occupation number $\langle n_{\mathbf{K}} \rangle$ in state \mathbf{K} , whereas the intensity $I(\omega - \Omega)$ of anti-Stokes peaks is $I(\omega - \Omega) \propto \langle n_{\mathbf{K}} \rangle$, corresponding to the destruction of one phonon. The ratio between Stokes and anti-Stokes intensities is thus often used to infer a measurement of temperature according to the relation $I(\omega + \Omega)/I(\omega - \Omega) \propto \langle n_{\mathbf{K}} \rangle / \langle n_{\mathbf{K}} + 1 \rangle = e^{-\hbar\Omega/(k_b T)}$.

Standard Raman scattering is a very weak phenomenon, as it can be covered by Rayleigh elastic scattering and fluorescence radiation. For this reason, a variant of Raman spectroscopy is often employed, that is named Resonance Raman (RR), able to enhance the signal intensity by several, in the range between 2 to 6, orders of magnitude. In RR, the wavelength of the exciting light source, such as e.g. a laser, coincides with the wavelength of an electronic transition. A schematic comparison

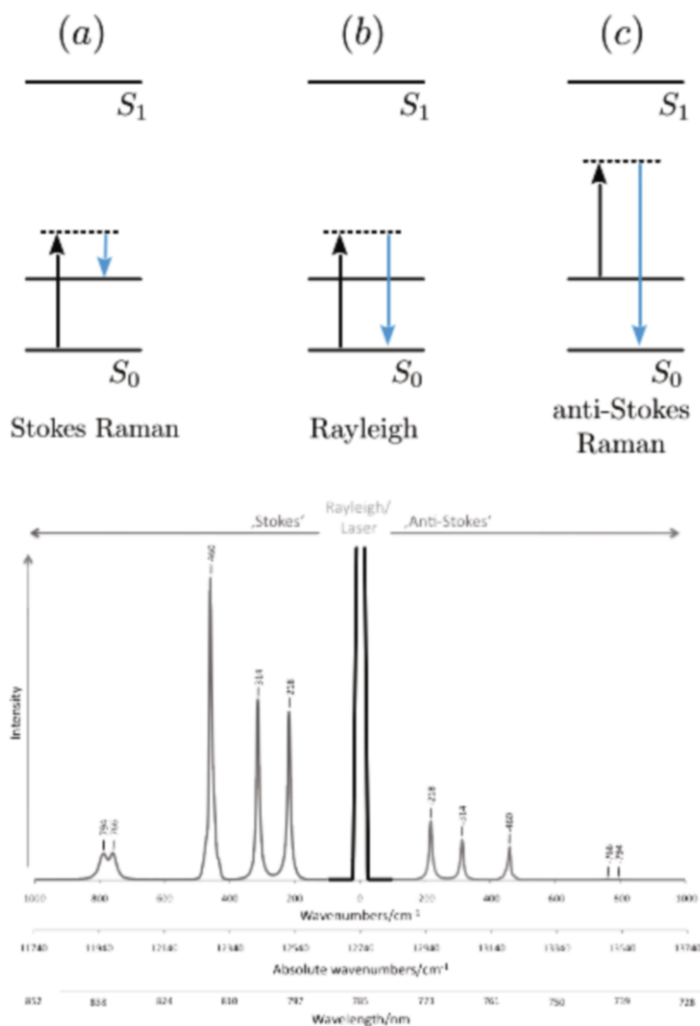


Fig. 3.12 Raman Spectroscopy. Top panel: schematics of the light scattering process. Inelastic Stokes (a) and anti-Stokes (c) transition. The elastic Rayleigh transition is shown in (b) for comparison (see text). Bottom panel: sketch of a Raman spectrum, displaying Stokes, Rayleigh elastic, and anti-stokes peaks

of the type of transitions involved in IR, Raman, RR, and Fluorescence Spectroscopies is illustrated in Fig. 3.13. One more additional variant of Raman techniques includes the coupling with a microscope to investigate very small samples or selecting characteristics of part of a sample with respect to the surroundings: nano-Raman or Tip-Enhanced Raman Spectroscopy are the general names for such techniques.

Selection Rules for IR and Raman Spectroscopies. As discussed so far, a system is IR-active if its dipole momentum μ changes, so that the intensity of an infrared

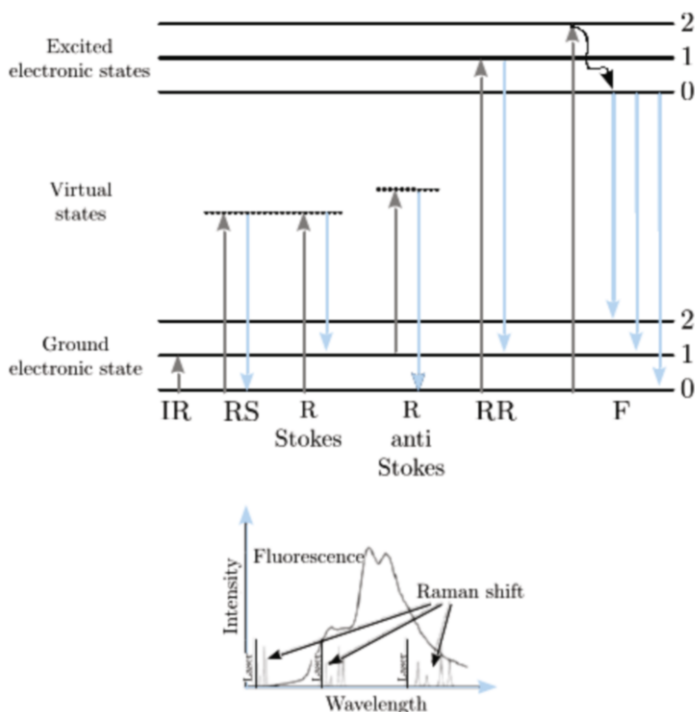


Fig. 3.13 Comparison among different spectroscopies: Infrared (IR), (elastic) Rayleigh Scattering (RS), Stokes and anti-Stokes Raman (R), Resonant Raman (RR) and Fluorescence (F). Energies are in cm^{-1} , that is in wavenumber $k = 2\pi/\lambda$ units according to the relation $E = hc/\lambda$ in terms of the light wavelength λ . Inset: sketch illustrating that fluorescence usually covers Raman signals

absorption peak $I_{IR} \propto (\partial\mu/\partial q)^2$ depends on the change of the dipole moment μ during the vibration. For a system to be Raman-active, the system polarizability α has to change instead and no permanent dipole moment is required for Raman scattering to occur. Therefore, as a rule of thumb when an inversion center of symmetry is present, vibrations that are symmetric with respect to the inversion center are Raman-active and IR-inactive, whereas molecular vibrations that are antisymmetric Raman-inactive are IR-active. This is known as the rule of mutual exclusion. In general, vibrational modes may exist that are both IR and Raman active. Just for illustration purposes one easy example of the complementarity between IR and Raman spectroscopies is displayed in Fig. 3.14 for gaseous carbon dioxide CO_2 molecules. Notice that the IR absorption peaks displayed on the top panel spot the asymmetric modes, such as the off-phase stretching modes at 2360 and 2324 cm^{-1} or the bending mode at 669 cm^{-1} , or else asymmetric combinations of fundamental modes. The Raman spectrum on the bottom spots instead the symmetric in-phase stretching mode: this is expected at 1332 cm^{-1} , though it is split into two peaks at 1385 cm^{-1} and 1278 cm^{-1} as it is almost resonant with the second overtone of the bending mode at 669 cm^{-1} .

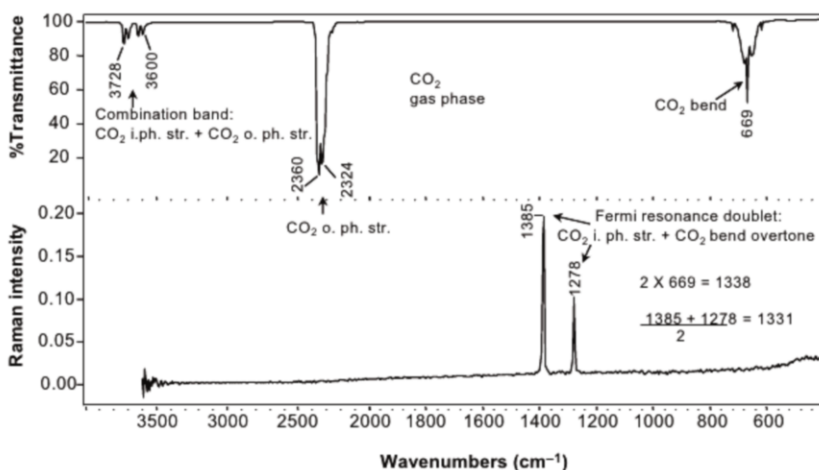


Fig. 3.14 IR and Raman spectra of carbon dioxide, showing the possibility of extracting complementary information. Top panel: IR absorption intensity in terms of % fraction of transmitted light. Bottom panel: Raman scattering intensity. Energies are in cm^{-1} . Discussion in the text [16]

Therefore, though some vibrational states may be both IR and Raman active, IR and Raman spectroscopies most often provide complementary information and are to be both used for a complete system analysis. Raman is especially useful to extract information on symmetric vibrational modes of non-polar compounds, in particular involving covalent bonds. IR spectroscopy is instead useful for asymmetric vibrational modes of polar compounds, in particular involving ionic bonds. The complementarity of the two methods shows up in technical manoeuvrability: at variance with IR, Raman spectroscopy does not require elaborated sample preparation but needs by far more expensive instrumentation.

Summary: concepts, tools, and procedures to know

Concepts and Tools

- **A body subjected to various kinds of stresses responds with various kinds of deformations.**
- In the elastic limit, the body responds to small acting stresses with a proportional change of either its volume or shape.
- The response is different to different kinds of stresses: linear compressions or stretching defines the Young modulus, volume compression or stretching the compressibility, shear stress the shear modulus.

- The relative stretching or compression along different unit vectors of a material are all related to each other.
- These macroscopic parameters characterizing the volume or shape elastic response of the material are related to the coefficients c_{ij} determined by internal forces between atoms, so that knowledge of the former infer information on the latter and viceversa.
- A distinction exists between volume and shape elasticity, this being found also in the dynamical behavior of crystal ions.
- **Mechanical material properties and vibrational wave propagation are related to each other.**
- The dynamical behavior of elementary portions of a continuous medium corresponds to (mechanical) wave propagation.
- The classical collective dynamics of a system of N ions moving around their equilibrium lattice positions can be described by the Hamiltonian of a set of independent harmonic oscillators.
- **In quantum physics, the collective harmonic oscillator variables are quantized to yield the phonons, that are quanta of elementary vibrational excitations.**
- Different boundary conditions determine different expressions for the allowed vibrational phases ϕ_v and thus normal mode frequencies.
- The frequencies of vibrational excitations of a finite linear chain of N identical atoms are dispersed into a band. For two atoms of different species, two phonon bands appear, and $2N$ vibrational frequencies. Gaps occur at the zone boundary, as for the case of electronic excitations.
- The frequencies of vibrational excitations also depend on the number of atoms. The dependence washes away while evolving from finite to infinite system and become independent of boundary conditions, also because the interactions between ions are screened by electrons and thus short-range in nature.
- Phonons can be either acoustical or optical.
- A crystal with r atoms in the base has $3r$ vibrational bands, three of them being acoustical and the remaining $3r - 3$ optical. In the long-wavelength acoustical modes, all the ions displace by the same amount independently of the species, while in the optical modes the ions move in each cell so to keep the center of mass be at rest.
- For the acoustical bands, the long-wavelength behavior of vibrational frequencies is linear in k , as atom displacements at long distances are very similar to each other, and a vanishingly small energy is thus needed to excite the system.
- **A phonon can be viewed as a quasiparticle describing a collective vibration of the atoms in the crystal.** A quasiparticle is a fictitious particle, with assigned momentum $\hbar\mathbf{k}$ and energy $\hbar\omega$, useful to describe the excitations of a system. At variance with true ones, a quasiparticle can be created and destroyed.
- **The density of vibrational (electronic) states determines the weight with which a given mode contributes to build up the average quantity that is observable.**

- Periodicity of $\omega_s(\mathbf{k})$ on the whole reciprocal lattice implies that in the first BZ $\omega_s(\mathbf{k})$ have at least one minimum, two saddle points, and one maximum.
- The type of critical points and the dimensionality of the system under study are reflected by characteristic and distinguished behaviors of the density of states with respect to frequency ω .
- The harmonic vibrational modes contribution to specific heat is the sum of the vibrational energies weighted by their average occupation at the given temperature, according to Bose-Einstein quantum statistics.
- **Finite response of a crystal to mechanical and thermal perturbations, measured e.g. by compressibility or thermal expansion and thermal conductivity, results from anharmonic terms, the response being vanishing within the bare harmonic approximation.**
- The (high-temperature) specific heat of classical harmonic solids is constant (Dulong and Petit law). The (low-temperature) specific heat of quantum harmonic solids vanishes as T^3 on the scale of the Debye temperature Θ_D embodying vibrational information characteristic of the crystal.
- The presence of significant anharmonic effects marks the behavior or even the existence of thermal properties of materials.
- **Determination of phonon spectra can be obtained by means of mechanical-wave probes, or X-ray inelastic scattering though from bright synchrotron sources, or inelastic neutron scattering, or else by light infrared absorption or Raman inelastic scattering processes.**
- Measurements of energy loss of incoming neutrons after they undergo inelastic scattering against vibrating nuclei, generally manifest as a background with better defined peaks on top of it due to first-order transitions.
- A molecule with same atoms cannot be IR-active, while molecules with different types of atoms are usually IR-active, even if the dipole moment is induced.
- A vibrational mode is Raman active if the structure polarizability changes during vibration.
- Raman and IR spectroscopies are often complementary: Raman spectroscopy is especially useful to extract information on symmetric vibrational modes of non-polar, in particular covalent, compounds, IR spectroscopy for asymmetric vibrational modes of polar, especially ionic, compounds. Variants of Raman, such as Resonance Raman or nano-Raman Tip-Enhanced Raman Spectroscopy is used for special needs.

Procedures

- Determination of system response coefficients determining the deformations driven by the action of mechanical stresses with given symmetry.
 - Calculation of vibrational frequencies and time-dependent displacements for a 1D chain, both finite and infinite.
 - Quantization of normal modes and calculation of phonon density of states.
 - Determination of system response coefficients determining mechanical or transport behavior driven by the action of thermal stresses.
-

3.13 Appendix. Kinetic and potential energies of ions in harmonic approximation

The kinetic and potential energy of ions in harmonic approximation is here calculated. The explicit calculation of the kinetic energy term is as follows:

$$\begin{aligned}
 K &= \sum_J \frac{1}{2} M_J \dot{s}_J^2 = \frac{1}{2} \sum_J M \dot{q}_J^2 \\
 &= \frac{M}{2} \sum_J \sum_{\alpha} \dot{a}_{\alpha}(t) Q_{J,\alpha} \sum_{\alpha'} \dot{a}_{\alpha'}(t) Q_{J,\alpha'} \\
 &= \frac{M}{2} \sum_{\alpha,\alpha'} \dot{a}_{\alpha} \dot{a}_{\alpha'} \delta_{\alpha,\alpha'} = \frac{M}{2} \sum_{\alpha} \dot{a}_{\alpha}^2.
 \end{aligned} \tag{3.175}$$

The above derivation states that a quadratic form with respect to the displacements can be transformed into a quadratic form into the collective variables a_{α} . The explicit calculation of the potential energy proceeds instead as follows:

$$\begin{aligned}
 \Phi &= \frac{1}{2} \sum_{J,K} \phi_{J,K} s_J s_K = \frac{M}{2} \sum_{J,K} D_{J,K} q_J q_K = \frac{M}{2} \sum_{J,K} D_{J,K} \sum_{\alpha} a_{\alpha}(t) Q_{J,\alpha} \sum_{\alpha'} a_{\alpha'}(t) Q_{K,\alpha'} \\
 &= \frac{M}{2} \sum_{\alpha,\alpha'} a_{\alpha} a_{\alpha'} \sum_J Q_{J,\alpha} \sum_K D_{J,K} Q_{K,\alpha'} = \frac{M}{2} \sum_{\alpha,\alpha'} a_{\alpha} a_{\alpha'} \sum_J Q_{J,\alpha} \omega_{\alpha'}^2 Q_{J,\alpha'} \\
 &= \frac{M}{2} \sum_{\alpha,\alpha'} a_{\alpha} a_{\alpha'} \omega_{\alpha'}^2 \delta_{\alpha,\alpha'} = \frac{M}{2} \sum_{\alpha} \omega_{\alpha}^2 a_{\alpha}^2.
 \end{aligned} \tag{3.176}$$

3.14 Appendix. Calculation of normalization coefficients

The normalization coefficient P of normal modes is here calculated, defined by the relation $P^2 \sum_{n=1}^N Q_n^2 = 1$, with $\varphi = v\pi/N$.

3.14.1 Case of one atom per cell

Since

$$\begin{aligned}
 Q_n^2 &= [A \sin(n\varphi) + B \cos(n\varphi)]^2 = A^2 \sin^2(n\varphi) + B^2 \cos^2(n\varphi) + AB \sin(2n\varphi) \\
 &= \frac{A^2}{2} [1 - \cos(2n\varphi)] + \frac{B^2}{2} [1 + \cos(2n\varphi)] + AB \sin(2n\varphi) \\
 &= \frac{A^2 + B^2}{2} + \frac{B^2 - A^2}{2} \cos(2n\varphi) + AB \sin(2n\varphi),
 \end{aligned} \tag{3.177}$$

one has that

$$P^2 \sum_{n=1}^N Q_n^2 = \frac{A^2 + B^2}{2} N, \quad (3.178)$$

provided that

$$\sum_{n=1}^N \cos(2n\varphi) = \sum_{n=1}^N \sin(2n\varphi) = 0. \quad (3.179)$$

On the other hand, since $\cos(2n\varphi) = (e^{2in\varphi} + e^{-2in\varphi})/2$, one has:

$$\begin{aligned} \sum_{n=1}^N e^{2in\varphi} &= e^{2i\varphi} \frac{1 - e^{2iN\varphi}}{1 - e^{2i\varphi}}, \\ \sum_{n=1}^N e^{-2in\varphi} &= e^{-2i\varphi} \frac{1 - e^{-2iN\varphi}}{1 - e^{-2i\varphi}}. \end{aligned} \quad (3.180)$$

Using $\varphi = v\pi/N$, one therefore obtains

$$\begin{aligned} \sum_{n=1}^N (e^{2in\varphi} + e^{-2in\varphi}) &= \frac{\cos(2\varphi) - 1 - \cos[2(N+1)\varphi] + \cos[2N\varphi]}{1 - \cos(2\varphi)} \\ &= \frac{\cos(2\varphi) - \cos[2(N+1)\varphi]}{1 - \cos(2\varphi)} \\ &= -2 \frac{\sin[(N+2)\varphi] \sin[N\varphi]}{1 - \cos(2\varphi)} = 0. \end{aligned} \quad (3.181)$$

In similar manner, one calculates that $\sum_{n=1}^N \sin(2n\varphi) = 0$.

3.14.2 Case of two atoms per cell

In this case it has been derived that the boundary conditions are satisfied when $\varphi_v = v\pi/(2N+1)$. The normalization condition for the displacements Q_n is now expressed by

$$A^2 \sum_{n=0}^{N-1} \sin^2[(2n+1)\varphi_v] + B^2 \sum_{n=1}^N \sin^2[2n\varphi_v] = 1. \quad (3.182)$$

Since

$$\sum_{n=0}^{N-1} \sin^2[(2n+1)\varphi_v] = \sum_{n=1}^N \sin^2[2n\varphi_v] = \frac{2N+1}{4},$$

from the set of equation (3.77) one finds that:

$$B_{v,s} = \frac{2\kappa - M_1 \omega_s^2(v)}{2\kappa \cos \phi_v} A_{v,s},$$

$$\sqrt{\frac{4}{2N+1}} = A_{v,s} \left[1 + \left(\frac{2\kappa - M_1 \omega_s^2(v)}{2\kappa \cos \phi_v} \right)^2 \right]^{\frac{1}{2}}. \quad (3.183)$$

3.15 Appendix. Bloch theorem

Consider (3.89) as the starting vector, from which a new vector is constructed so that the displacement of atom in cell h is mapped into the displacement of the atom in cell 0, the displacement of atom in cell $h+1$ into that of atom in cell 1, and so on. Since $Q_0 = Q_N$ and $Q_1 = Q_{N+1}$ and D_{m-n} depends only on the difference $m-n$, the new sought vector satisfies the set of equations (3.88) with the same eigenvalue. In the case of a non-degenerate eigenvalue, one has to have:

$$Q_{m+h} = c(h)Q_m, \quad \forall m. \quad (3.184)$$

The normalization condition for the two vectors leads to $|c(h)|^2 = 1$. Setting

$$Q_{m+1} = cQ_m, \quad |c|^2 = 1, \quad (3.185)$$

repeated applications of the property above lead to $c(h) = c^h$. Expressing

$$c = e^{ika} \quad (3.186)$$

with a the lattice period and k a real number, one finally obtains that

$$c(h) = e^{ihka}. \quad (3.187)$$

The relevant result is thus found, that the displacements of the atoms in whatever cell can be obtained from those of the atom in cell 0 by just applying an appropriate simple phase shift.

In the case of a degenerate eigenvalue of (3.88), the procedure is slightly modified. Consider e.g. a twofold degeneracy: therefore two linear independent eigenvalues are first selected, that is

$$\begin{array}{cc} Q'_0 & Q''_0 \\ \vdots & \vdots \\ Q'_{N-1} & Q''_{N-1}, \end{array} \quad (3.188)$$

from which two new vectors are constructed as before by means of a similar translation operation (3.184) applied to their components. The new vectors satisfy the set

of equations (3.88) with the same eigenvalue. In particular, in the case of a one-step translation one has

$$\begin{aligned}\underline{Q}'_{m+1} &= S_{1,1}\underline{Q}'_m + S_{1,2}\underline{Q}''_m, \\ \underline{Q}''_{m+1} &= S_{2,1}\underline{Q}'_m + S_{2,2}\underline{Q}''_m, \forall m.\end{aligned}\quad (3.189)$$

The matrix S with elements $S_{1,1}$, $S_{1,2}$, $S_{2,1}$ and $S_{2,2}$ is independent of m . Once again, norm conservation makes the transformation unitary. Consider now that the base vectors \underline{Q}' and \underline{Q}'' be modified by means of an m -independent transformation Γ such that

$$\begin{aligned}\underline{Q}'_m &= \Gamma_{1,1}\underline{Q}'_m + \Gamma_{1,2}\underline{Q}''_m \\ \underline{Q}''_m &= \Gamma_{2,1}\underline{Q}'_m + \Gamma_{2,2}\underline{Q}''_m \quad \forall m.\end{aligned}\quad (3.190)$$

Labeling with \underline{Q}'_m and \underline{Q}_m the two-component vectors

$$\begin{pmatrix} \underline{Q}'_m \\ \underline{Q}''_m \end{pmatrix}, \quad \underline{Q}'_m, \quad \underline{Q}''_m, \quad (3.191)$$

one finally has

$$\underline{Q}_{m+1} = \Gamma \underline{Q}_{m+1} = \Gamma S \underline{Q}_m = \Gamma S \Gamma^{-1} \underline{Q}_m. \quad (3.192)$$

One can thus revert back to the non-degenerate case, by choosing Γ such that $\Gamma S \Gamma^{-1}$ be diagonal, as it is always possible since S is a unitary transformation. Then, in all cases the relation (3.90) holds. The formal proof can be extended to the case of multiply-degenerate eigenvalues, by choosing the corresponding number of independent eigenvectors.

3.16 Appendix. Calculation of free energy, entropy, and total vibrational energy

Phonons are bosons. Reminding from quantum statistical mechanics the expression for the partition function Z of a system of bosons, for phonons one has:

$$Z = \sum_{n_{\mathbf{k}_1, s_1}, n_{\mathbf{k}_2, s_2}} e^{-\sum_{\mathbf{k}, s} \hbar \omega_s(\mathbf{k}) (n_{\mathbf{k}}(s) + \frac{1}{2}) \beta} = \prod_{\mathbf{k}, s} \sum_{n_{\mathbf{k}, s}} e^{-\hbar \omega_s(\mathbf{k}) (n_{\mathbf{k}}(s) + \frac{1}{2}) \beta} = \prod_{\mathbf{k}, s} \frac{e^{-\frac{\hbar \omega_s(\mathbf{k})}{2} \beta}}{1 - e^{-\hbar \omega_s(\mathbf{k}) \beta}}, \quad (3.193)$$

with $\beta \equiv (k_b T)^{-1}$. Notice that a chemical potential normally appears in the terms of the type $e^{[\beta(\hbar \omega - \mu)]}$, resulting from the conservation of the number of particles. This is not the case for a phonon system because of anharmonic effects, altering the conservation of the number of phonons.

Given the partition function, all the thermodynamic quantities follow. The vibrational free energy is

$$F_{ph} = -\frac{1}{\beta} \log Z = \sum_{\mathbf{k},s} \frac{\hbar \omega_s(\mathbf{k})}{2} + \frac{1}{\beta} \sum_{\mathbf{k},s} \log \left[1 - e^{-\hbar \omega_s(\mathbf{k})\beta} \right]. \quad (3.194)$$

To this, the free energy of the ions fixed at their equilibrium positions is to be added. Thus, the total free energy is:

$$F = F_0 + F_{ph}. \quad (3.195)$$

F_{ph} is a function of temperature and, in principle, of volume as well via $\omega_s(\mathbf{k})$. In fact, F_0 is only a function of volume.

It is interesting to notice that the frequencies $\omega_s(\mathbf{k})$ do not depend on volume, while within harmonic approximation. This result can be mathematically derived, demonstrating that the coefficients of the quadratic terms in a quadratic form with linear terms, are left unchanged by a reduction of the original form into a purely quadratic one. This is explicitly elaborated in Problem 3.10. The mathematical procedure represents real situations in crystals: when these are uniformly expanded, their translation vectors are shifted by a factor $(1 + \varepsilon)$ and therefore the lattice position $\mathbf{R}_m = m_1 \mathbf{a}_1 + m_2 \mathbf{a}_2 + m_3 \mathbf{a}_3$ is transformed into $\mathbf{R}'_m = (1 + \varepsilon) \mathbf{R}_m$, implying that the equilibrium positions are displaced. Under these conditions, the quadratic form referred to the original reference system includes linear terms, since forces are the non-vanishing spatial derivatives of the potential energy. If the coefficients of the quadratic terms are left unchanged, therefore the frequencies of the normal modes do not change in a displaced harmonic oscillator. Modifications are possible only due to anharmonic terms.

Reverting back to the thermodynamic quantities in harmonic approximation, the solid entropy $S = -(\frac{\partial F}{\partial T})_V$ and the internal energy $U = F + TS$ are finally given by

$$S = -k_B \sum_{\mathbf{k},s} \left[\log(1 - e^{-\hbar \omega_s(\mathbf{k})\beta}) - \hbar \omega_s(\mathbf{k})\beta \frac{1}{e^{\hbar \omega_s(\mathbf{k})\beta} - 1} \right],$$

$$U = U_0 + \sum_{\mathbf{k},s} \frac{\hbar \omega_s(\mathbf{k})}{2} + \sum_{\mathbf{k},s} \frac{\hbar \omega_s(\mathbf{k})}{e^{\hbar \omega_s(\mathbf{k})\beta} - 1}. \quad (3.196)$$

3.17 Appendix. Band calculation of the vibrational modes of crystals with one atom per cell

Consider from the main text the $3N$ equations of motion for the case of crystals with one atom per cell:

$$M \ddot{s}_{m,i} = - \sum_{\underline{n},l} \phi_{m-n,i}^{0,l} s_{n,l}. \quad (3.197)$$

These become

$$\ddot{s}_{m,i} = \sum_{n,l} D_{m-n,i}^{0,l} q_{n,l}, \quad (3.198)$$

with $D_{m,i}^{0,l} = \phi_{m,i}^{0,l}/M$. The Bloch theorem holds, conveniently expressed in the form

$$s_{m,i} = Q_i e^{i\mathbf{k} \cdot \mathbf{R}_m} e^{-i\omega t}, \quad (3.199)$$

where Q_i is the displacement of the given cell atom in direction i . Expression (3.198) leads to:

$$-\omega^2 Q_i = \sum_{h,l} D_{h,i}^{0,l} e^{i\mathbf{k} \cdot \mathbf{R}_h} Q_l. \quad (3.200)$$

Setting

$$t_{i,l}(\mathbf{k}) = \sum_h D_{h,i}^{0,l} e^{i\mathbf{k} \cdot \mathbf{R}_h}, \quad (3.201)$$

the set of $3N$ equations is reduced to a set of 3 equations

$$-\omega^2 Q_i = \sum_l t_{i,l}(\mathbf{k}) Q_l. \quad (3.202)$$

Since the Bravais lattice is symmetric under inversion operations, one has $\phi_{h,i}^{0,l} = \phi_{-h,i}^{0,l}$ and therefore

$$t_{i,l}(\mathbf{k}) = \sum_h D_{h,i}^{0,l} \cos(\mathbf{k} \cdot \mathbf{R}_h), \quad (3.203)$$

with $h \equiv (h_1, h_2, h_3)$ a set of integers. The eigenfrequencies are obtained after imposing that the determinant of the coefficient matrix be zero, that yields three solutions in ω^2 . Now, the properties $\omega_s(\mathbf{k}) = \omega_s(-\mathbf{k})$ and $\omega_s(\mathbf{k} + \mathbf{G}) = \omega_s(\mathbf{k})$ still hold and one can see that all the solutions are acoustical-like. Indeed, in the $\mathbf{k} \rightarrow 0$ limit (3.203) implies

$$t_{i,l}(\mathbf{k}) = \sum_h D_{h,i}^{0,l} \frac{(\mathbf{k} \cdot \mathbf{R}_h)^2}{2} = k^2 \sum_h D_{h,i}^{0,l} \frac{(\hat{\mathbf{k}} \cdot \mathbf{R}_h)^2}{2} = k^2 z_{i,l}(\hat{\mathbf{k}}), \quad (3.204)$$

with $\hat{\mathbf{k}} = \mathbf{k}/k$. The determinant takes the form

$$\begin{bmatrix} \frac{\omega^2}{k^2} - z_{1,1} & z_{1,2} & z_{1,3} \\ z_{2,1} & \frac{\omega^2}{k^2} - z_{2,2} & z_{2,3} \\ z_{3,1} & z_{3,2} & \frac{\omega^2}{k^2} - z_{3,3} \end{bmatrix}. \quad (3.205)$$

The three solutions have the form $\omega_s^2/k^2 = f_s(\hat{\mathbf{k}})$: this expresses the fact that the speed with which the deformation propagates in direction $\hat{\mathbf{k}}$ is $c_s(\hat{\mathbf{k}}) = \sqrt{f_s(\hat{\mathbf{k}})}$. The solutions $Q_i(\mathbf{k}, s)$ in turn satisfy the relations

$$\begin{aligned}\sum_i Q_i(\mathbf{k}, s) Q_i(\mathbf{k}, s') &= \delta_{s, s'}, \quad \text{orthogonality relation} \\ \sum_s Q_i(\mathbf{k}, s) Q_j(\mathbf{k}, s) &= \delta_{i, j}, \quad \text{completeness relation.}\end{aligned}\quad (3.206)$$

Periodic boundary conditions imply the existence of N allowed vectors \mathbf{k} in the Brillouin zone. Thus, the three branches yield $3N$ solutions, and the displacements can be expressed in terms of the normal variables $a_{s, \mathbf{k}}(t)$ as

$$\mathbf{s}_m(t) = \sum_{\mathbf{k}, s} a_{s, \mathbf{k}}(t) \mathbf{e}_s(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{R}_m}, \quad (3.207)$$

where $\mathbf{e}_s(\mathbf{k})$ represents the displacement of the atom in the cell located at the origin, that is a polarization vector with components $Q_i(\mathbf{k}, s)$.

These results can be extended to the case of a primitive cell with a base of r atoms with two differences. First, $3r$ phonon bands exist, three of them being acoustic with $\omega_s(\mathbf{k} \rightarrow 0) \simeq c_s(\hat{\mathbf{k}})k$ for $i = 1, 2, 3$, and $3r - 3$ optical with the long-wavelength displacements of the base atoms leaving the center of mass at rest. Second, the displacement of atom μ in the cell \mathbf{R}_m is given by

$$s_{m, \mu}(t) = \sum_{\mathbf{k}, s} a_{s, \mathbf{k}}(t) \sqrt{\frac{M}{M_\mu}} \mathbf{e}_{s, \mu}(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{R}_m}. \quad (3.208)$$

3.18 Appendix. Transition probabilities to first and second perturbative order

In order to simplify the calculation of the physical quantities of interest, consider a system described by an unperturbed Hamiltonian H_0 to which a time-dependent perturbation $\bar{H}(t)$ is added. Label the eigenstates of H_0 by ψ_i , with $i = 1, 2, 3$. One has to solve the time-dependent Schrödinger equation

$$\hbar \frac{\partial \psi}{\partial t} = (H_0 + \bar{H}(t)) \psi, \quad (3.209)$$

$$\bar{H}(t) = \begin{cases} 0 & \text{if } t < 0, \\ H'(t) & \text{if } 0 < t < \tau, \\ 0 & \text{if } t > \tau. \end{cases} \quad (3.210)$$

To make the notation short, the space and spin variables are not explicitly indicated in the wavefunctions. H_0 can in principle be the Hamiltonian of either a single or many particles. The general solution of (3.209) is:

$$\psi(t) = \sum_i \alpha_i(t) \psi_i e^{-i\varepsilon_i t / \hbar}, \quad (3.211)$$

with ε_i being the eigenvalues of the unperturbed Hamiltonian. The unknowns are now the functions $\alpha_i(t)$, which embody the system time dependence. For $t < 0$ and $t > \tau$, the wave function (3.211) is solution of (3.209) if the $\alpha_i(t)$ do not depend on time. In the interval $0 < t < \tau$, the quantities $\alpha_i(t)$ are to be chosen so that (3.209) is satisfied. One can also assume that for $t < 0$ $\alpha_1(0) = 1$, and $\alpha_i(0) = 0$ for $i = 2$ and 3. This implies that for $t < 0$ the system is found in the stationary state ψ_1 .

Once the quantities $\alpha_i(t)$ are calculated in the interval $0 < t < \tau$, one finds that the state

$$\psi(t) = \sum_i \alpha_i(\tau) \psi_i e^{-i\varepsilon_i t/\hbar} \quad (3.212)$$

is still a solution of (3.209) for $t > \tau$. The quantity $|\alpha_1(\tau)|^2$ represents the probability that at time τ the system be still in state ψ_1 , whereas $|\alpha_i(\tau)|^2$ with $i = 2, 3$ yields the probability that the system be in state ψ_i . The quantity of interest in physical applications is the transition probability per unit time $P_{12} = |\alpha_2(\tau)|^2/\tau$. Consider the simple case of a system with only two levels. Substitution of (3.211) into (3.209) followed by projection onto the states ψ_i , leads to the set of equations

$$\begin{aligned} i\hbar\dot{\alpha}_1(t) &= H'_{11}(t)\alpha_1(t) + H'_{12}(t)\alpha_2(t)e^{-i\omega_{21}t}, \\ i\hbar\dot{\alpha}_2(t) &= H'_{21}(t)\alpha_1(t)e^{i\omega_{21}t} + H'_{22}(t)\alpha_2(t), \end{aligned} \quad (3.213)$$

with $H'_{ij}(t) = \langle \psi_i | H'(t) | \psi_j \rangle$ and $\omega_{21} = (\varepsilon_2 - \varepsilon_1)/\hbar$. In most cases the matrix elements $H'_{ii}(t)$ are zero. If not, it is always possible to perform a transformation of variables that does the job [12] of setting those diagonal terms to zero. The following equations are thus conveniently considered:

$$\begin{aligned} i\hbar\dot{\alpha}_1(t) &= H'_{12}(t)\alpha_2(t)e^{-i\omega_{21}t}, \\ i\hbar\dot{\alpha}_2(t) &= H'_{21}(t)\alpha_1(t)e^{i\omega_{21}t}. \end{aligned} \quad (3.214)$$

The starting Hamiltonian is hermitian, so that $H'_{12}(t)e^{-i\omega_{21}t} = (H'_{21}(t)e^{i\omega_{21}t})^*$. Whenever these quantities do not depend on time, the set of (3.214) is easily integrable. Indeed, performing the derivative of the first equation and substituting the obtained expression in the second, one obtains

$$-\hbar^2\ddot{\alpha}_1(t) = |F|^2\alpha_1(t), \quad (3.215)$$

with $|F| = |H'_{21}(t)e^{i\omega_{21}t}|$. Equation (3.215) represents a harmonic oscillator with frequency $\bar{\omega} = |F|/\hbar$ and solutions $\alpha_1(t) \propto \cos(\bar{\omega}t)$ and $\alpha_2(t) \propto \sin(\bar{\omega}t)$. In fact, the system oscillates between the states ψ_1 and ψ_2 with period $T = 2\pi/\bar{\omega} = 2\pi\hbar/|F|$, $\bar{\omega}$ being named Rabi frequency.

Besides the above special case, the solution to the set of equations (3.214) or else to that with three states

$$\begin{aligned} i\hbar\dot{\alpha}_1(t) &= H'_{12}(t)\alpha_2(t)e^{-i\omega_{21}t} + H'_{13}(t)\alpha_3(t)e^{-i\omega_{31}t}, \\ i\hbar\dot{\alpha}_2(t) &= H'_{21}(t)\alpha_1(t)e^{i\omega_{21}t} + H'_{23}(t)\alpha_3(t)e^{-i\omega_{32}t}, \\ i\hbar\dot{\alpha}_3(t) &= H'_{31}(t)\alpha_1(t)e^{i\omega_{31}t} + H'_{32}(t)\alpha_2(t)e^{i\omega_{32}t}, \end{aligned} \quad (3.216)$$

is found by an iterative procedure as follows.

1. Write the relations

$$\begin{aligned}\alpha_1(t) &= 1 + \frac{1}{i\hbar} \int_0^t dt' \left[H'_{12}(t') \alpha_2(t') e^{-i\omega_{21}t'} + H'_{13}(t') \alpha_3(t') e^{-i\omega_{31}t'} \right], \\ \alpha_2(t) &= \frac{1}{i\hbar} \int_0^t dt' \left[H'_{21}(t') \alpha_1(t') e^{i\omega_{21}t'} + H'_{23}(t') \alpha_3(t') e^{-i\omega_{32}t'} \right], \\ \alpha_3(t) &= \frac{1}{i\hbar} \int_0^t dt' \left[H'_{31}(t') \alpha_1(t') e^{i\omega_{31}t'} + H'_{32}(t') \alpha_2(t') e^{i\omega_{32}t'} \right].\end{aligned}\quad (3.217)$$

These are equivalent to the equations in the set (3.216) and automatically include the initial conditions on the $\alpha_i(t)$.

2. Set the values of initial conditions on the right-hand-sides of the equations above. One thus obtains:

$$\begin{aligned}\alpha_1^{(1)}(t) &= 1, \\ \alpha_2^{(1)}(t) &= \frac{1}{i\hbar} \int_0^t dt' H'_{21}(t') e^{i\omega_{21}t'}, \\ \alpha_3^{(1)}(t) &= \frac{1}{i\hbar} \int_0^t dt' H'_{31}(t') e^{i\omega_{31}t'}.\end{aligned}\quad (3.218)$$

Here $|\alpha_2^{(1)}(\tau)|^2$ and $|\alpha_3^{(1)}(\tau)|^2$ represent the probabilities that at time τ the system have performed a transition from state ψ_1 to state ψ_2 and ψ_3 , respectively. These results are the content of first-order perturbation theory.

3. Once inserted back into (3.217), the values $\alpha_i^{(1)}(t)$ yield the results to second order in perturbation theory. The procedure can be further iterated along the same lines.
4. As a result, a succession of functions $\alpha_i^{(p)}(t)$ is constructed which, under very general conditions, converge to the set of integral equations (3.217) while $p \rightarrow \infty$. To second order for $\alpha_2^{(2)}(t)$, one thus obtains

$$\alpha_2^{(2)}(t) = \frac{1}{i\hbar} \int_0^t dt' H'_{21}(t') e^{i\omega_{21}t'} + \frac{1}{(i\hbar)^2} \int_0^t dt' H'_{23}(t') e^{-i\omega_{32}t'} \int_0^{t'} dt'' H'_{31}(t'') e^{i\omega_{31}t''}.\quad (3.219)$$

The relation above states that to first order the transition probability amplitude requires only the calculation of the matrix element of the perturbation between the initial and final states, whereas to second order the probability amplitude between states ψ_1 and ψ_2 depend on the intermediate matrix elements H'_{31} and H'_{23} . If a

system with more than three states were considered, the previous expression would be substituted by

$$\begin{aligned}\alpha_2^{(2)}(t) &= \frac{1}{i\hbar} \int_0^t dt' H'_{21}(t') e^{i\omega_{21}t'} \\ &+ \frac{1}{(i\hbar)^2} \sum_{j \neq 1} \int_0^t dt' H'_{2j}(t') e^{-i\omega_{j2}t'} \times \\ &\int_0^{t'} dt'' H'_{j1}(t'') e^{i\omega_{j1}t''}.\end{aligned}\quad (3.220)$$

In many cases the form $H'_{ij}(t)$ can be written as $H'_{ij}(t) = W_{ij} (e^{i\omega t} + e^{-i\omega t})$. For $\alpha_2^{(1)}(\tau)$ one thus obtains:

$$\alpha_2^{(1)}(\tau) = \frac{1}{i\hbar} W_{21} \left(\frac{e^{[i(\omega + \omega_{21})\tau]} - 1}{i(\omega + \omega_{21})} + \frac{e^{[i(-\omega + \omega_{21})\tau]} - 1}{i(-\omega + \omega_{21})} \right), \quad (3.221)$$

and $|\alpha_2^{(1)}(\tau)|^2$ results to be

$$|\alpha_2^{(1)}(\tau)|^2 = \frac{|W_{21}|^2}{\hbar^2} \left[\left| \frac{e^{[i(\omega + \omega_{21})\tau]} - 1}{i(\omega + \omega_{21})} \right|^2 + \left| \frac{e^{[i(-\omega + \omega_{21})\tau]} - 1}{i(-\omega + \omega_{21})} \right|^2 \right], \quad (3.222)$$

since in the $\tau \rightarrow \infty$ limit the interference terms are negligible.

The properties of the quantity

$$\left| \frac{e^{ix\tau} - 1}{ix} \right|^2 = 4 \frac{\sin^2(x\tau/2)}{x^2} \quad (3.223)$$

are as follows. First, the function (3.223) has a maximum at $x = 0$ with value τ^2 and width $2\pi/\tau$. Second, in the $\tau \rightarrow \infty$ limit the following relation holds:

$$4 \int_{-\infty}^{\infty} dx \frac{\sin^2\left(\frac{x\tau}{2}\right)}{x^2} = 2\pi\tau. \quad (3.224)$$

One then obtains for the transition probability per unit time P_{12}

$$P_{12} = \frac{2\pi}{\hbar^2} |W_{21}|^2 [\delta(\omega + \omega_{21}) + \delta(\omega - \omega_{21})]. \quad (3.225)$$

P_{12} is due to two processes: one transition corresponding to the condition $\varepsilon_2 - \varepsilon_1 + \hbar\omega = 0$, and a second transition with $\varepsilon_2 - \varepsilon_1 - \hbar\omega = 0$. If $\varepsilon_2 < \varepsilon_1$, only the first process may occur. If $\varepsilon_2 > \varepsilon_1$ only the second process takes place. The former case corresponds to the emission of a quantum with energy $\hbar\omega$, the final state having lower energy than the initial one. The latter corresponds to absorption of a quantum with energy $\hbar\omega$ and a final state with higher energy. One can thus conclude that the

interaction of the system with an external time-dependent field originates processes where energy quanta $\hbar\omega$ are either emitted or absorbed.

The calculation of the second-order term is performed after reverting back to (3.219). The transition probability per unit time, again contains one emission and one absorption term of the form:

$$\begin{aligned} P_{12}^{em} &= \frac{2\pi}{\hbar^4} \left| \frac{W_{13}W_{32}}{\omega + \omega_{31}} \right|^2 \delta(2\omega + \omega_{21}), \\ P_{12}^{as} &= \frac{2\pi}{\hbar^4} \left| \frac{W_{13}W_{32}}{-\omega + \omega_{31}} \right|^2 \delta(2\omega - \omega_{21}). \end{aligned} \quad (3.226)$$

If more than three states are involved, the corresponding expressions are:

$$\begin{aligned} P_{12}^{em} &= \frac{2\pi}{\hbar^4} \left| \sum_{j \neq 1} \frac{W_{1j}W_{j2}}{\omega + \omega_{j1}} \right|^2 \delta(2\omega + \omega_{21}), \\ P_{12}^{as} &= \frac{2\pi}{\hbar^4} \left| \sum_{j \neq 1} \frac{W_{1j}W_{j2}}{-\omega + \omega_{j1}} \right|^2 \delta(2\omega - \omega_{21}). \end{aligned} \quad (3.227)$$

If $H'_{12}(t) = W_{12} [e^{i\omega t} + e^{-i\omega t}]$, one can find an approximate solution to the system (3.214) as

$$\begin{aligned} i\hbar\dot{\alpha}_1(t) &= W_{12} \left[e^{i(\omega - \omega_{12})t} + e^{-i(\omega + \omega_{12})t} \right] \alpha_2(t), \\ i\hbar\dot{\alpha}_2(t) &= W_{21} \left[e^{i(\omega + \omega_{12})t} + e^{-i(\omega - \omega_{12})t} \right] \alpha_1(t). \end{aligned} \quad (3.228)$$

Consider the frequencies in the vicinity of ω_{12} . In this case, the fast oscillating terms can be neglected, that are those proportional to $e^{-i(\omega + \omega_{12})t}$. The following equations are thus considered:

$$\begin{aligned} i\hbar\dot{\alpha}_1(t) &= W_{12}\alpha_2(t)e^{i(\omega - \omega_{12})t}, \\ i\hbar\dot{\alpha}_2(t) &= W_{21}\alpha_1(t)e^{-i(\omega - \omega_{12})t}. \end{aligned} \quad (3.229)$$

Using the transformations $\alpha_1(t) = a_1(t)e^{i\Delta\omega t/2}$, $\alpha_2(t) = a_2(t)e^{(-i\Delta\omega t/2)}$, and $\Delta\omega = \omega - \omega_{12}$, one obtains

$$\begin{aligned} i\hbar \left(\dot{a}_1(t) + i\frac{\Delta\omega}{2}a_1(t) \right) &= W_{12}a_2(t), \\ i\hbar \left(\dot{a}_2(t) - i\frac{\Delta\omega}{2}a_2(t) \right) &= W_{21}a_1(t). \end{aligned} \quad (3.230)$$

The latter equations are solved after setting $a_1(t) = Ae^{i\lambda t}$ e $a_2(t) = Be^{i\lambda t}$. By substitution, the following set of linear homogeneous equations is found

$$-\hbar \left(\lambda + \frac{\Delta\omega}{2} \right) A - W_{12}B = 0,$$

$$-W_{21}A - \hbar \left(\lambda - \frac{\Delta\omega}{2} \right) B = 0. \quad (3.231)$$

This set admits a non trivial solution if and only if $\lambda^2 - (\Delta\omega/2)^2 - |W_{12}|^2/\hbar^2 = 0$, and provided that the ratio $B/A = -\hbar(\lambda + \Delta\omega/2)/W_{12}$. For $\alpha_1(t)$ e $\alpha_2(t)$ one thus obtains:

$$\begin{aligned} \alpha_1(t) &= \left(A_1 e^{i\lambda t} + A_2 e^{-i\lambda t} \right) e^{i\frac{\Delta\omega}{2}t}, \\ \alpha_2(t) &= -\frac{\hbar}{W_{12}} e^{-i\frac{\Delta\omega}{2}t} \times \\ &\quad \left[\left(\lambda + \frac{\Delta\omega}{2} \right) A_1 e^{i\lambda t} + \left(-\lambda + \frac{\Delta\omega}{2} \right) A_2 e^{-i\lambda t} \right]. \end{aligned} \quad (3.232)$$

Insertion of the initial conditions $\alpha_1(0) = 1$ e $\alpha_2(0) = 0$ yields

$$\begin{aligned} A_1 &= \frac{\lambda - \Delta\omega/2}{2\lambda}, \\ A_2 &= \frac{\lambda + \Delta\omega/2}{2\lambda}. \end{aligned} \quad (3.233)$$

Finally, one has

$$\begin{aligned} \alpha_1(t) &= \left(\frac{\lambda - \Delta\omega/2}{2\lambda} e^{i\lambda t} + \frac{\lambda + \Delta\omega/2}{2\lambda} e^{-i\lambda t} \right) e^{i\frac{\Delta\omega}{2}t}, \\ \alpha_2(t) &= -\frac{\hbar}{W_{12}} \frac{1}{2\lambda} e^{-i\frac{\Delta\omega}{2}t} \left[\left(\lambda^2 - \left(\frac{\Delta\omega}{2} \right)^2 \right) e^{i\lambda t} - \left(\lambda^2 - \left(\frac{\Delta\omega}{2} \right)^2 \right) e^{-i\lambda t} \right] \\ &= -\frac{i\hbar}{W_{12}\lambda} \left(\lambda^2 - \left(\frac{\Delta\omega}{2} \right)^2 \right) \sin(\lambda t) e^{-i\frac{\Delta\omega}{2}t}. \end{aligned} \quad (3.234)$$

The probability that at time t the system be found in state ψ_2 is:

$$\begin{aligned} |\alpha_2(t)|^2 &= \frac{\hbar^2}{|W_{12}|^2 \lambda^2} \left(\lambda^2 - \left(\frac{\Delta\omega}{2} \right)^2 \right)^2 \sin^2(\lambda t) \\ &= \frac{|W_{12}|^2 \sin^2(\lambda t)}{\hbar^2 \lambda^2}. \end{aligned} \quad (3.235)$$

For $W_{12} \rightarrow 0$ and in the $t \rightarrow \infty$ limit, this becomes a term of (3.225), since to lowest order $\lambda = (\omega - \omega_{12})/2$.

3.19 Appendix. Effective expression of the electron-phonon Hamiltonian H_{e-ph}

In order to derive the effective expression for H_{e-ph} let first decompose α into all the different indexes needed to identify an ion inside a crystal. The electron-phonon interaction Hamiltonian can be written as

$$\begin{aligned}
 H_{e-ph} &= - \sum_{m,i,\mu} \nabla_{\mathbf{r}_i} V_{\mu}(\mathbf{R}_{m,\mu} - \mathbf{r}_i) \cdot \mathbf{s}_{m,\mu} = \\
 &- \sum_{m,i,\mu} \nabla_{\mathbf{r}_i} V_{\mu}(\mathbf{R}_{m,\mu} - \mathbf{r}_i) \cdot \sum_{\mathbf{q},s} \sqrt{\frac{\hbar}{2\omega_s(\mathbf{q})}} (b_{s,-\mathbf{q}}^{\dagger} + b_{s,\mathbf{q}}) \mathbf{e}_{\mu}(\mathbf{q},s) \times \\
 &e^{i\mathbf{q} \cdot \mathbf{R}_m} \sqrt{\frac{1}{M_{\mu}}} = \sum_{i,\mathbf{q},s} \sqrt{\frac{\hbar}{2\omega_s(\mathbf{q})}} (b_{s,-\mathbf{q}}^{\dagger} + b_{s,\mathbf{q}}) \sum_{\mu} \sqrt{\frac{1}{M_{\mu}}} \mathbf{e}_{\mu}(\mathbf{q},s) \cdot \\
 &\sum_m \nabla_{\mathbf{r}_i} V_{\mu}(\mathbf{R}_{m,\mu} - \mathbf{r}_i) e^{i\mathbf{q} \cdot \mathbf{R}_m} \\
 &= \sum_{i,\mathbf{q},s} \sqrt{\frac{\hbar}{2\omega_s(\mathbf{q})}} (b_{s,-\mathbf{q}}^{\dagger} + b_{s,\mathbf{q}}) F(\mathbf{q},s,\mathbf{r}_i). \tag{3.236}
 \end{aligned}$$

Here,

$$F(\mathbf{q},s,\mathbf{r}_i) = \sum_{\mu} \sqrt{\frac{1}{M_{\mu}}} \mathbf{e}_{\mu}(\mathbf{q},s) \cdot \sum_m \nabla_{\mathbf{r}_i} V_{\mu}(\mathbf{R}_{m,\mu} - \mathbf{r}_i) e^{i\mathbf{q} \cdot \mathbf{R}_m}. \tag{3.237}$$

In particular, one has:

$$\begin{aligned}
 F(\mathbf{q},s,\mathbf{r}_i + \mathbf{R}_h) &= \sum_{\mu} \sqrt{\frac{1}{M_{\mu}}} \mathbf{e}_{\mu}(\mathbf{q},s) \cdot \\
 &\sum_m \nabla_{\mathbf{r}_i} V_{\mu}(\mathbf{R}_{m,\mu} - \mathbf{r}_i - \mathbf{R}_h) e^{i\mathbf{q} \cdot \mathbf{R}_m} \\
 &= \sum_{\mu} \sqrt{\frac{1}{M_{\mu}}} \mathbf{e}_{\mu}(\mathbf{q},s) \cdot \sum_{m'} \nabla_{\mathbf{r}_i} V_{\mu}(\mathbf{R}_{m',\mu} - \mathbf{r}_i) e^{i\mathbf{q} \cdot (\mathbf{R}_{m'} + \mathbf{R}_h)} \\
 &= e^{i\mathbf{q} \cdot \mathbf{R}_h} F(\mathbf{q},s,\mathbf{r}_i) \tag{3.238}
 \end{aligned}$$

3.19.1 Calculation of the matrix element $F(\mathbf{q},s,\mathbf{r})$

The function $F(\mathbf{q},s,\mathbf{r})$ is spin independent. Thus, the matrix element

$$\langle m_p \mathbf{k}_p \alpha_p | F(\mathbf{q},s,\mathbf{r}) | m'_p \mathbf{k}'_p \alpha_p \rangle$$

can be cast in the form

$$\begin{aligned}
 & \langle m_p \mathbf{k}_p | F(\mathbf{q}, s, \mathbf{r}) | m'_p \mathbf{k}'_p \rangle \\
 &= \int d\mathbf{r} \phi_{m_p}(\mathbf{k}_p, \mathbf{r})^* F(\mathbf{q}, s, \mathbf{r}) \phi_{m'_p}(\mathbf{k}'_p, \mathbf{r}) \\
 &= \sum_h \int_{V_c} d\mathbf{r} \phi_{m_p}(\mathbf{k}_p, \mathbf{r} + \mathbf{R}_h)^* F(\mathbf{q}, s, \mathbf{r} + \mathbf{R}_h) \phi_{m'_p}(\mathbf{k}'_p, \mathbf{r} + \mathbf{R}_h) \\
 &= \sum_h \int_{V_c} d\mathbf{r} e^{-i\mathbf{k}_p \cdot \mathbf{R}_h} \phi_{m_p}(\mathbf{k}_p, \mathbf{r})^* e^{i\mathbf{q} \cdot \mathbf{R}_h} F(\mathbf{q}, s, \mathbf{r}) e^{i\mathbf{k}'_p \cdot \mathbf{R}_h} \phi_{m'_p}(\mathbf{k}'_p, \mathbf{r}) \\
 &= \sum_h e^{i(\mathbf{q} + \mathbf{k}'_p - \mathbf{k}_p) \cdot \mathbf{R}_h} \int_{V_c} d\mathbf{r} \phi_{m_p}(\mathbf{k}_p, \mathbf{r})^* F(\mathbf{q}, s, \mathbf{r}) \phi_{m'_p}(\mathbf{k}'_p, \mathbf{r}), \quad (3.239)
 \end{aligned}$$

with $\phi_{m_p}(\mathbf{k}_p, \mathbf{r})$ and $\phi_{m'_p}(\mathbf{k}'_p, \mathbf{r})$ the Bloch functions. The factor given by the integral is non-zero if

$$\mathbf{q} + \mathbf{k}'_p - \mathbf{k}_p = \mathbf{G}, \quad (3.240)$$

yielding a selection rule.

Problems with solutions

3.1. Consider the situation in which the stiffness coefficients c_{ij} are modified by δc_{ij} . Show that the contact forces perform a work determining a variation δF of the free energy of the elementary cube. Calculate this variation and show that $c_{ij} = c_{ji}$.

Solution. Consider a cube with unit-length side, whose uniform deformation is described by the quantity e_{ij} . Consider the variations de_{ij} of these deformations. The contact forces on a face perform a work δL

$$\begin{aligned}
 \delta L &= \mathbf{F} \cdot d\mathbf{R} = (X\hat{\mathbf{x}} + Y\hat{\mathbf{y}} + Z\hat{\mathbf{z}}) \cdot [du(\mathbf{r})\hat{\mathbf{x}} + dv(\mathbf{r})\hat{\mathbf{y}} + dw(\mathbf{r})\hat{\mathbf{z}}] \\
 &= Xdu(\mathbf{r}) + Ydv(\mathbf{r}) + Zdw(\mathbf{r}), \quad (3.241)
 \end{aligned}$$

with

$$\begin{aligned}
 du(\mathbf{r}) &= x d\epsilon_{xx} + y d\epsilon_{yx} + z d\epsilon_{zx}, \\
 dv(\mathbf{r}) &= x d\epsilon_{xy} + y d\epsilon_{yy} + z d\epsilon_{zy}, \\
 dw(\mathbf{r}) &= x d\epsilon_{xz} + y d\epsilon_{yz} + z d\epsilon_{zz}, \quad (3.242)
 \end{aligned}$$

given by (3.2). Consider a cube around the origin with dimensions dx , dy and dz . The work done on the surface perpendicular to the x -axis at $dx/2$ is given by

$$\delta L_1 = [X_x(0, 0, 0)du + Y_x(0, 0, 0)dv + Z_x(0, 0, 0)dw] dydz$$

where $du = d\epsilon_{xx}dx/2$, $dv = d\epsilon_{xy}dx/2$ and $dw = d\epsilon_{xz}dx/2$. At the surface perpendicular to the x -axis at $-dx/2$ it is given by

$$\delta L_2 = -[X_x(0, 0, 0)(-du) + Y_x(0, 0, 0)(-dv) + Z_x(0, 0, 0)(-dw)] dydz.$$

To the first order the sum of the two terms gives

$$\delta L_1 + \delta L_2 = dxdydx [X_x d\epsilon_{xx} + Y_x d\epsilon_{xy} + Z_x d\epsilon_{xz}]$$

where $X_x = X_x(0,0,0)$, $Y_x = Y_x(0,0,0)$ and $Z_x = Z_x(0,0,0)$. For the other two pair of faces perpendicular to the y - and z -axis it is obtained

$$\begin{aligned} \delta L_3 + \delta L_4 &= dxdydx [X_y d\epsilon_{yx} + Y_y d\epsilon_{yy} + Z_y d\epsilon_{yz}], \\ \delta L_5 + \delta L_6 &= dxdydx [X_z d\epsilon_{zx} + Y_z d\epsilon_{zy} + Z_z d\epsilon_{zz}]. \end{aligned} \quad (3.243)$$

The total work per unit volume δL is given by the sum of the six calculated terms. One obtains:

$$\delta L = X_x de_{xx} + Y_y de_{yy} + Z_z de_{zz} + Y_z de_{yz} + Z_x de_{zx} + X_y de_{xy}.$$

δL can be identified with the variation of the density of free energy $d\Psi$, that is a state function, so that:

$$\begin{aligned} \frac{\partial \Psi}{\partial e_{xx}} &= X_x, \quad \frac{\partial \Psi}{\partial e_{yy}} = Y_y, \quad \frac{\partial \Psi}{\partial e_{zz}} = Z_z, \\ \frac{\partial \Psi}{\partial e_{yz}} &= Y_z, \quad \frac{\partial \Psi}{\partial e_{zx}} = Z_x, \quad \frac{\partial \Psi}{\partial e_{xy}} = X_y. \end{aligned}$$

Substitution of the constitutive relations (3.12) in the above equation, immediately yields that $c_{ij} = c_{ji}$.

3.2. Demonstrate that in a cubic system the only non-vanishing elastic constants are: $c_{11} = c_{22} = c_{33}$, $c_{44} = c_{55} = c_{66}$, $c_{12} = c_{13} = c_{23}$.

Solution. Consider a system with an operation transforming the axes (x, y, z) into (x', y', z') , so that a given physical quantity changes its expression while moving from the original to the final reference. The elastic coefficients in particular, would assume the values c'_{ij} . In the case of a proper or not proper rotation, the following relation exists between c'_{ij} and c_{ij} :

$$c'_{ij} = \sum_{k,l} n_{ijkl} c_{kl},$$

where the actual expression of the coefficients n_{ijkl} depend on the transformation. In the case of the identical transformation, the previous equation gets the form

$$c_{ij} = \sum_{k,l} n_{ijkl} c_{kl}.$$

Specification of coefficients n_{ijkl} determines whether selected c_{ij} have the same value in the two reference systems or whether they are zero.

Consider the rotational symmetry operations that transform a cube in itself. Among them are the rotations by $2\pi/3$ around the cube diagonals. As a consequence of such operation, the following axes mapping occurs: $x \rightarrow x' \equiv y$, $y \rightarrow y' \equiv z$, $z \rightarrow$

$z' \equiv x$. Therefore, one obtains $c_{11} = c'_{11} = c_{22}$, $c_{22} = c'_{22} = c_{33}$ with $c_{11} = c_{22} = c_{33}$, $c_{12} = c'_{12} = c_{23}$, $c_{23} = c'_{23} = c_{31}$ with $c_{12} = c_{23} = c_{31}$, and finally $c_{44} = c'_{44} = c_{55}$ and $c_{55} = c'_{55} = c_{66}$ with $c_{44} = c_{55} = c_{66}$. Besides, the cube is invariant after rotations by π around axes x , y , and z . Thus, a rotation around the z -axis produces $x \rightarrow x' \equiv -x$, $y \rightarrow y' \equiv -y$, $z \rightarrow z' \equiv z$ so that the coefficient $c_{14} = c'_{14} = -c_{14}$, implying $c_{14} = 0$. Proceeding in similar manners for the remaining symmetry operations, one finds that $c_{14} = c_{15} = c_{26} = c_{24} = c_{25} = c_{36} = c_{34} = c_{35} = c_{16} = c_{46} = c_{54} = c_{56} = 0$.

3.3. Consider a cubic system with the condition $c_{11} - c_{12} = 2c_{44}$. Show how this ensures that longitudinal and transverse waves are characterized respectively by the velocities $\sqrt{c_{11}/\rho}$ and $\sqrt{c_{44}/\rho}$, independently of the propagation direction.

Solution. The set of differential equations (3.30) for a system with cubic symmetry is transformed into the set of equations (3.32) for an isotropic material, provided that $c_{11} - c_{12} = 2c_{44}$ and $c_{44} = \mu$, $c_{12} = \lambda$. The velocity of transverse waves is thus $\sqrt{\mu/\rho}$, while that of longitudinal waves is $\sqrt{(\lambda + 2\mu)/\rho}$, as demonstrated in the main text. Actually, one can check that the particular solutions of the linear set of equations (3.35) along the three different directions $(1, 0, 0)$, $(1, 1, 0)$ and $(1, 1, 1)$, and with transverse-wave velocities

$$\sqrt{c_{44}/\rho}, \quad \sqrt{(c_{11} - c_{12})/(2\rho)}, \quad \sqrt{(c_{11} - c_{12} + c_{44})/(3\rho)}$$

and longitudinal-wave velocities

$$\sqrt{c_{11}/\rho}, \quad \sqrt{(c_{11} + c_{12} + 2c_{44})/(2\rho)}, \quad \sqrt{(c_{11} + 2c_{12} + 4c_{44})/(3\rho)},$$

become $\sqrt{c_{44}/\rho}$ and $\sqrt{c_{11}/\rho}$.

3.4. Consider the set of equations (3.59), where additional forces $-\mathbf{F}$ are introduced in the first equation and \mathbf{F} in the third. Determine the stretching of the wire under the condition that the displacements be independent of time. Consider now M^2 wires under the above conditions. Arrange the wires to be parallel to each other so that they form a parallelepiped with sides Ma and height Na , and connect them by means of springs. Determine the Young modulus of such a system and show that it is reduced while the size Na increases.

Solution. Consider the indicated set of equation. One has: $\Delta s_2 = s_2 - s_1 = F/\kappa$, $\Delta s_3 = s_3 - s_2 = F/\kappa$, $\Delta s_N = s_N - s_{N-1} = F/\kappa$. Thus: $\Delta l = \Delta s_2 + \Delta s_3 + \dots \Delta s_N = (N - 2)F/\kappa$. Given that $l = Na$, one then obtains $\Delta l/l = (1 - 2/N)F/(\kappa a)$. In the 3D system described in the second part of the problem text, the force acting on a surface unit is $F_i/A = M^2 F/(M^2 a^2) = F/a^2$. Therefore $\Delta l/l = (1 - 2/N)aF/(\kappa a^2)$, leading to the Young modulus $G = \kappa/[a(1 - 2/N)]$. By inspection, larger G values correspond to smaller N .

3.5. Given the set of equations (3.60)

$$-M\omega^2 Q_1 = \kappa(Q_2 - Q_1), n = 1,$$

$$\begin{aligned} -M\omega^2 Q_n &= \kappa(Q_{n+1} - Q_n) + \kappa(Q_{n-1} - Q_n), n = 2, \dots, N-1, \\ -M\omega^2 Q_N &= \kappa(Q_{N-1} - Q_N). n = N, \end{aligned} \quad (3.244)$$

show that

1. the equations of the second line with $n = 2, 3, \dots, N-1$ are all satisfied by $Q_n = A \sin(n\varphi) + B \cos(n\varphi)$ with A, N , and φ arbitrary, provided that

$$\omega^2 = \frac{2\kappa}{M} (1 - \cos \varphi); \quad (3.245)$$

2. the first and last equation are satisfied provided that

$$\begin{aligned} -A \sin \varphi + B(1 - \cos \varphi) &= 0, \\ \alpha A + \beta B &= 0, \end{aligned} \quad (3.246)$$

with

$$\begin{aligned} \alpha &= (-1 + 2 \cos \varphi) \sin(N\varphi) - \sin[(N-1)\varphi], \\ \beta &= (-1 + 2 \cos \varphi) \cos(N\varphi) - \cos[(N-1)\varphi]; \end{aligned} \quad (3.247)$$

3. The determinant of the coefficients in the set of equations (3.246) vanishes provided that $\sin(N\varphi) = 0$.

Solution.

1. Substituting Q_n in the equations of the second line of the set of equations (3.60), one obtains that it is satisfied if (3.245) is verified whatever the values of n, A, B , and φ be.
2. After a simple calculation, substitution of Q_n in the first of (3.244) and the use of (3.245) yields the first line of (3.246). Finally the substitution of Q_n in the last equation of the set of equations (3.244) and the use of (3.245), gives the second line of (3.246).
3. The determinant of the coefficients matrix of the set of equations (3.246) is zero if $D = \beta \sin \varphi + \alpha (1 - \cos \varphi) = 0$. One finds:

$$\begin{aligned} D &= [(-1 + 2 \cos \varphi) \cos(N\varphi) - \cos(N-1)\varphi] \sin \varphi \\ &\quad + [(-1 + 2 \cos \varphi) \sin(N\varphi) - \sin(N-1)\varphi] (1 - \cos \varphi) \\ &= (-1 + 2 \cos \varphi) [\cos(N\varphi) \sin \varphi + \sin(N\varphi)(1 - \cos \varphi)] \\ &\quad - \cos(N-1)\varphi \sin \varphi - \sin(N-1)\varphi (1 - \cos \varphi) \\ &= (-1 + 2 \cos \varphi) [-\sin(N-1)\varphi + \sin(N\varphi)] + \sin(N-2)\varphi - \sin(N-1)\varphi \\ &= -\sin(N\varphi) + \sin(N-2)\varphi + 2 \cos \varphi [-\sin(N-1)\varphi + \sin(N\varphi)] \\ &= -2 \cos(N-1)\varphi \sin \varphi + 2 \cos \varphi \sin N\varphi - 2 \cos \varphi \sin(N-1)\varphi \\ &= -2 \sin N\varphi + 2 \cos \varphi \sin N\varphi. \end{aligned} \quad (3.248)$$

The last line above vanishes provided that $\sin(N\varphi) = 0$.

3.6. Determine the vibrational energy and specific heat of 1D and 2D systems, within the Debye model.

Solution. The system has only one branch with $\omega(k) = vk$ in the 1D case, and two branches with $\omega_1(k) = v_1k$ and $\omega_2(k) = v_2k$ in the 2D case. The number of modes in the two cases is, respectively:

$$\begin{aligned} dN &= \frac{L}{2\pi} dk = \frac{L}{2\pi} \frac{dk}{d\omega} d\omega = \frac{L}{2\pi} \frac{1}{v} d\omega \quad (1D); \\ dN &= \frac{S}{(2\pi)^2} d\mathbf{k} = \sum_{i=1}^2 \frac{S}{(2\pi)^2} 2\pi k \frac{1}{v_i} d\omega \\ &= \frac{S}{2\pi} \omega \left(\frac{1}{v_1^2} + \frac{1}{v_2^2} \right) d\omega \quad (2D). \end{aligned} \quad (3.249)$$

Here, the following conditions are to be satisfied

$$\begin{aligned} N &= \frac{L}{2\pi} \frac{1}{v} \omega_D \quad (1D); \\ 2N &= \frac{S}{2\pi} \left(\frac{1}{v_1^2} + \frac{1}{v_2^2} \right) \frac{\omega_D^2}{2} \quad (2D), \end{aligned} \quad (3.250)$$

where N is number of atoms and ω_D the Debye frequency. The vibrational energy thus results to be:

$$\begin{aligned} E &= \frac{L}{2\pi} \frac{1}{v} \int_0^{\omega_D} d\omega \frac{\hbar\omega}{e^{\beta\hbar\omega} - 1} \quad (1D); \\ E &= \frac{S}{2\pi} \left(\frac{1}{v_1^2} + \frac{1}{v_2^2} \right) \int_0^{\omega_D} d\omega \omega \frac{\hbar\omega}{e^{\beta\hbar\omega} - 1} \quad (2D), \end{aligned} \quad (3.251)$$

with $\beta = (k_b T)^{-1}$. Setting $x = \beta\hbar\omega$, one has $d\omega = dx/(\beta\hbar)$ and the energies can be expressed, respectively, as:

$$\begin{aligned} E &= \frac{L}{2\pi} \frac{1}{v\hbar\beta^2} \int_0^{\beta\hbar\omega_D} dx \frac{x}{e^x - 1} \quad (1D); \\ E &= \frac{S}{2\pi} \left(\frac{1}{v_1^2} + \frac{1}{v_2^2} \right) \frac{1}{\hbar^2\beta^3} \int_0^{\beta\hbar\omega_D} dx \frac{x^2}{e^x - 1} \quad (2D). \end{aligned} \quad (3.252)$$

If $k_b T \ll \hbar\omega_D$, the energies scale as T^2 and T^3 , respectively. The corresponding specific heats per particle are, in turn:

$$\begin{aligned} c_V &= k_b \frac{k_b T}{\hbar\omega_D} \left[2 \int_0^{\beta\hbar\omega_D} dx \frac{x}{e^x - 1} - \left(\frac{\hbar\omega_D}{k_b T} \right)^2 \frac{1}{e^{\beta\hbar\omega_D} - 1} \right], \\ c_V &= 4k_b \left(\frac{k_b T}{\hbar\omega_D} \right)^2 \left[3 \int_0^{\beta\hbar\omega_D} dx \frac{x^2}{e^x - 1} - \left(\frac{\hbar\omega_D}{k_b T} \right)^3 \frac{1}{e^{\beta\hbar\omega_D} - 1} \right]. \end{aligned} \quad (3.253)$$

In the low temperature limit $k_b T \ll \hbar \omega_D$, the specific heats scale as T and T^2 . In the high-temperature range, the Dulong-Petit regime is found with $c_V = k_b$ and $c_V = 2k_b$, respectively.

3.7. Determine the dispersion relations for a linear chain with two atoms per cell and alternating κ and κ' stiffness values.

Solution. Begin with the equations of motion

$$\begin{aligned} M_1 \ddot{s}_{n,1} &= \kappa(s_{n,2} - s_{n,1}) + \kappa'(s_{n-1,2} - s_{n,1}) \\ M_2 \ddot{s}_{n,2} &= \kappa'(s_{n+1,1} - s_{n,2}) + \kappa(s_{n,1} - s_{n,2}). \end{aligned} \quad (3.254)$$

Using the Bloch theorem and the usual dependence on time, one can write:

$$s_{n,i} = Q_i e^{i(kna - \omega t)}.$$

The two equations to be considered are

$$\begin{aligned} (\kappa + \kappa' - M_1 \omega^2) Q_1 - (\kappa + \kappa' e^{-ika}) Q_2 &= 0 \\ -(\kappa + \kappa' e^{ika}) Q_1 + (\kappa + \kappa' - M_2 \omega^2) Q_2 &= 0. \end{aligned} \quad (3.255)$$

These originate the two branches, represented by:

$$\omega^2 = \frac{MF \pm \sqrt{M^2 F^2 - 8\kappa\kappa' M_1 M_2 (1 - \cos(ka))}}{2M_1 M_2} = \frac{MF \pm \sqrt{\Delta}}{2M_1 M_2},$$

with $M \equiv M_1 + M_2$ and $F \equiv \kappa + \kappa'$.

It is easy and interesting the study of the special cases: a) κ or $\kappa' = 0$; b) $\kappa \gg \kappa'$.

3.8. Consider the compressibility B of an alkali-halogen material, that is $B = -V(\partial P / \partial V)$, with P the pressure and V the volume, and r_0 be the equilibrium distance of the first-neighbors ions with opposite sign. Let also the interaction between the ions be of the form:

$$\phi_{ij} = \frac{\pm e^2}{r_{ij}} + \frac{\gamma}{r_{ij}^n}.$$

Show that the conditions $P = 0$ and $\partial U / \partial r = 0$, with U the potential energy of each ion, are equivalent. Determine γ and n . Apply these results to materials crystallizing in the structure of NaCl and ClCs.

Solution. Let $2N$ be the number of atoms and r the distance between first neighbors. The potential energy of each ion in the field of all the others is:

$$U = -\frac{e^2}{r} \sum_{j \neq 0} \left(\pm \frac{1}{p_{0,j}} \right) + \frac{\gamma}{r^n} \sum_{j \neq 0} \frac{1}{p_{0,j}^n},$$

where a plus or minus sign is taken according to whether the j -th neighbors have opposite or equal charge with respect to that of the given reference ion, and $p_{0,j} =$

r_{0j}/r is its corresponding distance scaled in units of r . Since $\alpha_m = \sum_{j \neq 0} (\pm 1/p_{0,j})$ is the Madelung constant, the total energy can be expressed as:

$$U = N \left(-\frac{e^2 \alpha_m}{r} + \frac{c}{r^n} \right),$$

with $c = \sum_{j \neq 0} \gamma/p_{0,j}^n$. In the conventional cubic cell there are q ions. Since the side a of the conventional cubic cell is αr , the system volume is $V = 2N(\alpha r)^3/q = N\beta r^3$ or else $r = V^{1/3}/(\beta N)^{1/3}$, with $\beta = \frac{2\alpha^3}{q}$. The pressure is:

$$\begin{aligned} P &= -\frac{\partial U}{\partial V} = -\frac{\partial U}{\partial r} \frac{\partial r}{\partial V} \\ &= -\frac{1}{3\beta r^2} \left[\frac{\alpha_m e^2}{r^2} - \frac{nc}{r^{n+1}} \right]. \end{aligned} \quad (3.256)$$

The condition $P = 0$ implies that

$$\frac{e^2 \alpha_m}{r_0^2} - n \frac{c}{r_0^{n+1}} = 0,$$

that is the same equation obtained after setting $\partial U/(\partial r) = 0$. The compressibility is

$$\begin{aligned} B &= -V \frac{\partial P}{\partial V} = -V \frac{\partial P}{\partial r} \frac{\partial r}{\partial V} \\ &= \frac{V}{3\beta} \frac{1}{3} \frac{V^{-2/3}}{(\beta N)^{1/3}} \left[\left(-4 \frac{e^2 \alpha_m}{r^5} + n(n+3) \frac{c}{r^{n+4}} \right) \right]. \end{aligned} \quad (3.257)$$

At equilibrium one has

$$\begin{aligned} n \frac{c}{r_0^n} &= \frac{e^2 \alpha_m}{r_0} \\ B &= \frac{r_0}{9\beta} \frac{1}{r_0^4} \left[-4 \frac{e^2 \alpha_m}{r_0} + \frac{n(n+3)c}{r_0^n} \right], \end{aligned} \quad (3.258)$$

and therefore

$$\begin{aligned} B &= \frac{1}{9\beta r_0^3} (n-1) \phi_{el}, \\ \phi_{el} &= \frac{e^2 \alpha_m}{r_0}. \end{aligned} \quad (3.259)$$

Since ϕ_{el} , r_0 and B are given, the above equation yields the remaining unknowns

$$n = 1 + \frac{9\beta B r_0^3}{\phi_{el}},$$

whereas

$$c = \frac{\phi_{el} r_0^n}{n}.$$

Quantitative analysis of the various structures can thus be carried out after reminding that in the case of NaCl-like structures one has $q = 8$ and $\alpha = \beta = 2$. In the case of CsCl-like structures, one has $q = 2$, $\alpha = 2/\sqrt{3}$, and $\beta = 8/(3\sqrt{3})$.

3.9. Consider an infinite linear atomic chain with elastic interactions limited to first neighbors. Consider the case where the site located at the origin is occupied by one atom of the same mass but stiffness constant different from that of the remaining atoms in the chain. Show that a mode may exist, with solution $s_n(t) = s_0 e^{[-k(\omega)|n|a - i\omega t]}$.

Solution. The equations of motion for the atoms are:

$$\begin{aligned} m\ddot{s}_n &= \kappa(s_{n-1} - 2s_n + s_{n+1}), \quad n \neq 0 \\ m\ddot{s}_0 &= \kappa'(s_{-1} - 2s_0 + s_1), \quad n = 0. \end{aligned} \quad (3.260)$$

Insertion of s_n with $n > 0$ yields:

$$-m\omega^2 = \kappa [e^{ka} - 2 + e^{-ka}].$$

In the case with $n < 0$, one finds the same expression. For $n = 0$, one has:

$$-m\omega^2 = \kappa' [e^{-ka} - 2 + e^{ka}]. \quad (3.261)$$

The two last expressions provide ka and ω . Their side-by-side ratio yields

$$\frac{\kappa}{\kappa'} = \frac{2(1 - e^{-ka})}{2 - e^{-ka} - e^{ka}}.$$

Setting $x = e^{-ka}$, the previous equation becomes

$$\left(2 - \frac{\kappa}{\kappa'}\right)x^2 - 2x\left(1 - \frac{\kappa}{\kappa'}\right) - \frac{\kappa}{\kappa'} = 0,$$

whose solutions are

$$x = \frac{1}{\left(2 - \frac{\kappa}{\kappa'}\right)} \left[1 - \frac{\kappa}{\kappa'} \pm 1\right].$$

The solution with plus sign yields $x = 1$ and $ka = 0$, that is to be discarded on physical grounds. The solution with minus sign yields

$$x = -\frac{\kappa}{\kappa'} \frac{1}{2 - \frac{\kappa}{\kappa'}},$$

that leads to the determination of ka by means of

$$ka = \ln \left(1 - \frac{2\kappa'}{\kappa} \right).$$

From (3.261) the expression for the frequency is obtained, that is:

$$\omega^2 = \frac{4\kappa'}{m} \frac{1}{2 - \frac{\kappa}{\kappa'}}.$$

This is positive whenever $\kappa < 2\kappa'$. It thus results that $1 - 2\kappa'/\kappa < 0$ and therefore one has to write

$$ka = i\pi + \log \left(\frac{2\kappa'}{\kappa} - 1 \right).$$

As a result, one has the following cases: $\kappa' < \kappa < 2\kappa'$ implying that $\Re(ka) < 0$, and $\kappa < \kappa'$ implying $\Re(ka) > 0$. The solution in the former case is not acceptable, since ω is real but s_n is exponentially increasing with increasing n . The solution in the latter case is acceptable, since ω is real and s_n exponentially decreasing with n .

Notice that in a perfect lattice with one atom per cell, the phonon band has its maximum value with frequency $\omega_m^2 = 4\kappa/m$. By inspection, one can check that $4\kappa/m < 4\kappa'/[m(2 - \kappa/\kappa')]$, and thus the new mode lies outside the band.

3.10. Consider the quadratic form with linear terms

$$f = ax^2 + by^2 + 2cxy + dx + fy.$$

Show that when transformed into a purely quadratic form, it leaves the coefficients a , b , and c unchanged.

Solution. Determine first the stationary points of the quadratic form. To this aim, one imposes the conditions:

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2ax + 2cy + d = 0, \\ \frac{\partial f}{\partial y} &= 2by + 2cx + f = 0, \end{aligned} \tag{3.262}$$

leading to one single point (x_0, y_0) if $ab - c^2 \neq 0$. Under these conditions, the following variable transformation is conveniently operated:

$$\begin{aligned} x &= x_0 + x', \\ y &= y_0 + y'. \end{aligned} \tag{3.263}$$

The form f thus becomes:

$$\begin{aligned}
 f &= a(x_0 + x')^2 + b(y_0 + y')^2 + 2c(x_0 + x')(y_0 + y') + d(x_0 + x') + f(y_0 + y') \\
 &= ax_0^2 + by_0^2 + 2cx_0y_0 + dx_0 + fy_0 + x'(2ax_0 + 2cy_0 + d) \\
 &\quad + y'(2by_0 + 2cx_0 + f) + ax'^2 + by'^2 + 2cx'y' \\
 &= ax_0^2 + by_0^2 + 2cx_0y_0 + dx_0 + fy_0 + ax'^2 + by'^2 + 2cx'y'. \tag{3.264}
 \end{aligned}$$

The above results demonstrate that the quadratic form with the presence of linear terms can be transformed into a purely quadratic form (except for a constant), and that the coefficients a , b , and c are left unchanged.

3.11. Consider the 2D band of a square lattice represented by the expression $E(k)/E_0 = 2 - \cos k_x a - \cos k_y a$. Calculate the density of states in the vicinity of points $\Gamma \equiv (0, 0)$, $X \equiv (\frac{\pi}{a}, 0)$, and $L \equiv (\frac{\pi}{a}, \frac{\pi}{a})$ of the BZ.

Solution. The calculation of the density-of-states integral requires first the identification of the constant energy lines. It is first convenient to set $\varepsilon(k) = E(k)/E_0$ and consider that $\nabla \varepsilon(k) \equiv (a \sin k_x a, a \sin k_y a)$. In the vicinity of point Γ , one finds that no lines with constant energy may exist whenever $\varepsilon < 0$, whereas if $\varepsilon > 0$ the line of constant energy around Γ is a circumference with radius $R = \sqrt{2\varepsilon}/a$. Using these information, the density-of-states integral can be easily calculated to be:

$$\begin{aligned}
 g(E) &= \frac{2S}{(2\pi)^2} \int_{\varepsilon(k)=\varepsilon} \frac{dl}{|\nabla E(k)|} \\
 &= \frac{2S}{(2\pi)^2} \left[2\pi k \frac{1}{a(k_x^2 + k_y^2)^{\frac{1}{2}}} \right]_{k=R} = \frac{S}{\pi} \frac{1}{a}, \tag{3.265}
 \end{aligned}$$

with S being the sample surface and dl the infinitesimal line element at constant energy in the 2D band. The density of states in the vicinity of the remaining BZ points can be worked out along the same lines.

In the vicinity of point L one has $\varepsilon(k) = 2 + \cos x + \cos y$, with $k_x a = \pi - x$ and $k_y a = \pi - y$, and therefore $\varepsilon(k) = 4 - (x^2 + y^2)$. If $\varepsilon < 4$, the previous case is recovered. If $\varepsilon > 4$, no constant-energy lines exist.

In the vicinity of point X , the energy can be written as $\varepsilon(k) = 2 + \cos x - \cos y$, with $k_x a = \pi - x$ and $k_y a = y$. Therefore, one has $\varepsilon(k) = 2 - \frac{x^2}{2} + \frac{y^2}{2}$. The surface with constant energy is represented by $\varepsilon = 2 - \frac{x^2}{2} + \frac{y^2}{2}$. If $\varepsilon > 2$, one must have $y^2 > x^2$. If $\varepsilon < 2$ instead, one must have $x^2 > y^2$. In any event the two results are identical, due to the symmetry of the function. Since $\partial \varepsilon(k)/(\partial k_x) = -ax$ and $\partial \varepsilon(k)/(\partial k_y) = ay$, one finds that if $\varepsilon > 2$, $y = \sqrt{2\varepsilon + x^2}$ with $\bar{\varepsilon} = \varepsilon - 2$. The density of states is

$$g(E) = \frac{2S}{(2\pi)^2} \int_{\bar{\varepsilon} = -\frac{x^2}{2} + \frac{y^2}{2}} \frac{dl}{|\nabla \varepsilon(k)|}.$$

Since $dl = \sqrt{(dk_x)^2 + (dk_y)^2} = \frac{1}{a} \sqrt{(dx)^2 + (dy)^2}$, $|\nabla \varepsilon(k)| = a(x^2 + y^2)$, and the constant-energy line is characterized by the condition $d\bar{\varepsilon} = 0 = -x dx + y dy = 0$,

one has that $dy/dx = x/y$ and therefore:

$$\begin{aligned}
 g(E) &= \frac{2S}{(2\pi)^2} \frac{1}{a^2} \int dx \frac{\sqrt{1 + \left(\frac{dy}{dx}\right)^2}}{\sqrt{x^2 + y^2}} = \frac{2S}{(2\pi)^2} \frac{1}{a^2} \int_{\sqrt{2\bar{\epsilon}}}^R dx \frac{\sqrt{1 + \frac{x^2}{y^2}}}{\sqrt{x^2 + y^2}} \\
 &= \frac{2S}{(2\pi)^2} \log \left[\frac{R}{\sqrt{2\bar{\epsilon}}} + \sqrt{1 + \frac{R^2}{2\bar{\epsilon}}} \right] \\
 &= \frac{2S}{(2\pi)^2} \left[\bar{A} - \frac{1}{2} \log(\epsilon - 2) \right]. \tag{3.266}
 \end{aligned}$$

Here, $R \gg 1$ and $\bar{A} = \log(2R) - \log(2)/2$. A similar expression can be found in the case where $\epsilon < 2$.

Problems without solutions

3.12. Consider a cylindrical material of radius R_0 , length l and shape elasticity μ submitted to torque \mathbf{M} . Show that the torsion angle θ is $M = \mu\pi\theta R_0^4/l$.

3.13. Derive expression (3.27), $\mu = G/[2(1 + \sigma)]$.

3.14. Consider an elastic cubic material. Calculate the dispersion relation when the wavevector is $\mathbf{k} \equiv (k_x, k_y, 0)$ and the displacement has components $(u_0, v_0, 0)$.

3.15. Consider a system composed by three masses: two of them with equal mass M and the third with mass M_1 placed between the former two. The three masses interact via an elastic force with constant k . Calculate the frequency of the normal modes for this system.

3.16. Calculate the total vibrational energy and the specific heat of the system in Problem 3.15.

3.17. Calculate the band structure of an infinite linear chain characterized by elastic interactions with first and second neighbors.

3.18. Use (3.44) and write the components of the displacement vector \mathbf{D} .

References

1. L. Bloomfield: How Everything Works. Making physics out of ordinary. John Wiley and Sons, USA (2008)
2. F. Fumi: Appunti delle lezioni di Fisica dello Stato Solido. Università di Genova (1973)

3. R. Feynman, R.B. Leighton, and M. Sands: The Feynman lectures in Physics. 1st ed. Inter European Editions, Amsterdam (1975)
4. M.M.J. Treacy, T.W. Ebbesen, and J.M. Gibson, *Nature* **381**, 678 (1996)
5. P. Poncharal, Z.L. Wang, D. Ugarte, and A. de Heer, *Science* **283**, 1513 (1999)
6. G. Fasol, M. Tanaka, and H. Sakaki, *Phys. Rev. B* **38**, 6056 (1988)
7. G. Gilat and R.M. Nicklow, *Phys. Rev.* **143**, 487 (1966)
8. R. Stedman, L. Almqvist, and G. Nilsson, *Phys. Rev.* **162**, 549 (1967)
9. G. Grosso and G. Pastori Parravicini: *Solid State Physics*. 1st ed. Academic Press, New York (2000)
10. P. Giannozzi, S. De Gironcoli, P. Pavone, and S. Baroni, *Phys. Rev. B* **43**, 7231 (1991)
11. *American Institute of Physics Handbook* (third edition), edited by D.E. Gray. McGraw-Hill Book Company, New York (1972)
12. A.S. Davydov: *Quantum mechanics*. 1st ed. Pergamon Press, Oxford (1986)
13. K.A. Geshneider: *Solid state Physics*. Vol. 16. Academic Press, New York (1964)
14. Landolt-Burnstein Tabellen, N.S. III/1 Springer-Verlag. Berlin (1979); N.S. III/16, (1971); N.S. III/7 (1973)
15. P. Y. Yu and M. Cardona: *Fundamentals of Semiconductors*. Springer, Berlin (1996)
16. P. J. Larkin: *Infrared and Raman Spectroscopy. Principles and Spectral Interpretation*. Elsevier, Amsterdam (2011)

Chapter 4

Transport and Equilibrium Properties

Abstract The equations of motion for electrons and holes are derived within the semiclassical approximation for the treatment of external forces. The motion of electrons and holes in the presence of either electric or magnetic fields or both is investigated, with reference to their charge and spin properties. Conducting, semiconducting and insulating materials are characterized according to their transport behavior. The results obtained in this way are compared with the experimental data observing significant incompatibilities, so that scattering mechanisms of the carriers (electrons or holes) with phonons and impurities are introduced in order to explain the observations. Thus, electric and thermal conductivity are discussed with some length for metals and semiconductors, also referring to $p - n$ junctions. Then, selected characteristic transport properties of superconductors are briefly pointed out. Finally, electric conductivity in nanostructures is briefly discussed.

4.1 Introduction

Did you ever ask why efficient heaters must have large surfaces of well specified materials or why our body is so efficient in keeping our temperature at about 37°C , or else how our many electronic devices might work with progressively fast rates while being progressively miniaturized or how far can precision measurements be pushed. The content of this chapter may open up useful answers [1].

The first three Chapters have indeed been devoted to build up knowledge on the structure of the fixed lattice and of electron and phonon excitation energies. The present and the next Chapters are then aimed to describe the transport of the system when subjected to external disturbances. Here, the charge and energy current in response to external electric and magnetic fields and to temperature gradients are investigated, and the related transport coefficients are introduced. Frequency-dependent polarization and conductivity in response to external time-dependent electromagnetic fields are instead addressed in the next Chap. 5, where the opti-

cal properties of crystals and nanostructures are discussed in terms of dielectric and magnetic susceptibilities as well as optical conductivity.

In fact, the transport coefficients and static susceptibilities investigated in the present Chapter, can be obtained from the limiting behaviors of one and the same function describing the response of the system to external slowly varying space and time-dependent weak perturbations. This is a consequence of conservation laws and the content of the so-called Kubo relations. On the other hand, the response to time and space-dependent external fields investigated in the next Chapter actually provides a probe of the system structure, as already experienced at the end of each Chaps. 1-3 while discussing the experimental methods: this is the content of the so-called fluctuation and dissipation theorem. The connections among all these apparently different pieces of knowledge are outlined in Chap. 6 within the theoretical framework of the Density and Current-Density Functional Theory, suited to treat strongly correlated and inhomogeneous systems.

The first part of this Chapter is devoted to introduce the equations of motion for one single electron in a crystal in the presence of external forces due to electric and/or magnetic fields, along with the corresponding solution. The tools developed in Chap. 2 are used as ingredients to this aim, along with the quasi-classical treatment of external forces. Already within this simple approach, the difference between materials with metallic, semiconducting, or insulating behavior emerges and can be understood. The most relevant emerging concepts are those related to effective mass and cyclotron frequency. However, the theoretical predictions are unable to fit experimental data, until scattering mechanisms are introduced, acting among electrons and lattice vibrations or else impurities and disorder agents that break the crystal periodicity. In fact, the interaction between electrons and fixed ions is already embodied in the band-structure construction of electronic states and plays no role in transport phenomena. On the same footing, electron-electron interactions do not play a significant role even in metals, since scattering events are generally rare. Special conditions of reduced dimensionality and quantum degeneracy might show quite interesting exceptions, whose detailed discussion is however beyond the scope of this textbook.

The second part of this Chapter investigates the electronic contribution to a number of conduction phenomena: electric and thermal conductivity and thermoelectric effects in metals and semiconductors, and thermal conductivity in insulators. To this aim, the equilibrium thermodynamics of crystalline and nanometric systems are first usefully recalled, and then their modifications are explored in the presence of external fields. Within this description, the role played by the different quantum statistics naturally emerges. A detailed comparison between theory and experiments requires a microscopic treatment of scattering mechanisms. This is developed in a manner that is conceived to be as simple as possible, to keep the reading affordable. In the case of semiconductors, the doping mechanism by means of substitutional impurities of different species is especially relevant to understand even the simplest semiconductor-based device, that is a $p-n$ junction [2, 3].

The final part of this Chapter is devoted to applications of electric conduction in nanostructures inserted between metallic electrode plates. The assumption of ideal

zero resistance in the contact regions is taken here, for simplicity. The treatment is further simplified by neglecting the interactions with nanostructure lattice vibrations and admitting the presence of one or more electronic states that can be either occupied or empty. Though simplified, the present treatment leads to the basic determination of the current-voltage characteristic function between the two electrodes [4].

At variance with electronic and vibrational properties, the treatment of transport behavior turns out to be simpler in 3D crystalline than in nanometric structures. For this reason, in this Chapter and at variance with the previous ones, 3D idealized structures are discussed prior to finite structures.

4.2 Equations of motion for electrons in a crystal and in the presence of external fields

The dynamics of electrons in a crystal is a complex problem, first of all since the stationary states are described by Bloch functions $\psi_{n,\mathbf{k}}(\mathbf{r})$ and the crystal momentum $\hbar\mathbf{k}$ is not the electron momentum. Very often though, external fields act onto the system, which are slowly varying on the length scale of a lattice distance. In this case, a simplified treatment can be put forward, based on semiclassical approximation and quantum theory of electronic band states. As a starting step, one considers the time-dependent state:

$$\Psi(\mathbf{r}, t) = \sum_n \int_{V_r} d\mathbf{k} g_n(\mathbf{k}) \psi_{n,\mathbf{k}}(\mathbf{r}) e^{-iE_n(\mathbf{k})t/\hbar}, \quad (4.1)$$

that is expanded in the complete set of Bloch states $\psi_{n,\mathbf{k}}(\mathbf{r})$. In the following the more simple case is first considered, in which the spin degrees of freedom can be neglected, as in most situations occurs. By construction, this state contains contributions from all electronic bands, the weight of each band being $g_n(\mathbf{k})$. In a large number of cases, external fields significantly act to modify the electronic states only within the same band, and no transitions among different bands occur. From now on, this simplified and frequently occurring situation is considered, remarking though that the present treatment can be extended to the more complex case after introducing suited procedure modifications [5]. In this simplified case, the wavefunction (4.1) is written as

$$\Psi(\mathbf{r}, t) = \int_{V_r} d\mathbf{k} g_n(\mathbf{k}) \psi_{n,\mathbf{k}}(\mathbf{r}) e^{-iE_n(\mathbf{k})t/\hbar}. \quad (4.2)$$

The average wave number associated to the wavefunction is conveniently well defined if its average width $\Delta\mathbf{k}$ is much smaller than the size of the Brillouin Zone (BZ) volume. To this aim, the function $g_n(\mathbf{k})$ has to be significantly non-vanishing only on a small portion of the BZ, or else the average spatial width $\Delta\mathbf{r}$ of the state wavefunction (4.2) is to be much larger than the cell volume size in the direct lattice. If \mathbf{R} is a lattice vector, expression (4.2) along with the Bloch theorem implies that:

$$\begin{aligned}\Psi(\mathbf{r} + \mathbf{R}, t) &= \int_{V_r} d\mathbf{k} g_n(\mathbf{k}) \psi_{n,\mathbf{k}}(\mathbf{r} + \mathbf{R}) e^{-iE_n(\mathbf{k})t/\hbar} \\ &= \int_{V_r} d\mathbf{k} g_n(\mathbf{k}) \psi_{n,\mathbf{k}}(\mathbf{r}) e^{i[\mathbf{k} \cdot \mathbf{R} - E_n(\mathbf{k})t/\hbar]}.\end{aligned}\quad (4.3)$$

This expression represents a wavepacket in the variable \mathbf{R} with frequency $E_n(\mathbf{k})/\hbar$. But it does not precisely represent it, the reason being that \mathbf{R} is in fact a discrete variable. However, since $g_n(\mathbf{k})\psi_{n,\mathbf{k}}(\mathbf{r})$ is non vanishing in a small portion of the BZ, expression (4.3) represents a wavepacket of much larger size than the elementary cell size and thus \mathbf{R} can be considered a continuous variable. The condition for this approximation to hold can be expressed as

$$|\Delta \mathbf{R}| = \sqrt{\langle \Psi | \mathbf{R}^2 | \Psi \rangle - \langle \Psi | \mathbf{R} | \Psi \rangle^2} \gg a, \quad (4.4)$$

with a the lattice parameter. Thus

Concept

Whenever external potentials acting on the quantum particle are slowly varying over the $|\Delta \mathbf{R}|^3$ volume as in (4.4), the semiclassical approximation to the quantum-mechanical treatment can be adopted where the particle position is identified by the average

$$\bar{\mathbf{R}} = \langle \Psi | \mathbf{R} | \Psi \rangle, \quad (4.5)$$

and its velocity by the wavepacket group velocity

$$\mathbf{v}_{n,g} = \frac{1}{\hbar} \nabla E_n(\mathbf{k}). \quad (4.6)$$

Equations (4.5)-(4.6) represent the first useful tool to describe the dynamical behavior of electrons within this semiclassical approximation. A second equation is obtained after identifying the work done by the external force \mathbf{F}_{ext} on the electron, as the variation of its band energy. That is:

$$\frac{dE_n(\mathbf{k})}{dt} = \mathbf{F}_{\text{ext}} \cdot \mathbf{v}_{n,g}. \quad (4.7)$$

On the other hand one has

$$dE_n(\mathbf{k}) = \nabla E_n(\mathbf{k}) \cdot d\mathbf{k},$$

and, using (4.6) one obtains:

$$\mathbf{v}_{n,g} \cdot (\hbar d\mathbf{k} - \mathbf{F}_{\text{ext}} dt) = 0. \quad (4.8)$$

Equation (4.8) is valid if either the vector between braces vanishes or its direction is perpendicular to $\mathbf{v}_{n,g}$. In the former case, one has

$$\hbar \frac{d\mathbf{k}}{dt} = \mathbf{F}_{\text{ext}}, \quad (4.9)$$

corresponding to the situation in which a longitudinal external field, such as an electric field, is present. In the latter case, the external force on the right-hand side has to be complemented by a transverse term: this has indeed to be orthogonal to $\mathbf{v}_{n,g}$, and does not perform any work on the particle. This is the case of the presence of, e.g., a magnetic field, so that one has

$$\hbar \frac{d\mathbf{k}}{dt} = \mathbf{F}'_{\text{ext}} - \frac{e}{c} \mathbf{v}_{n,g} \wedge \mathbf{B}, \quad (4.10)$$

where now \mathbf{F}'_{ext} is the longitudinal (non-magnetic) part of the external force and $(e/c)\mathbf{v}_{n,g} \wedge \mathbf{B}$ is the transverse (magnetic) part of it. Notice that within this scheme, the periodic crystal potential determines the dynamical behavior of the particle only through the electronic band. In fact, (4.9) and (4.10) have the form of Newton equations, as if one had a particle with momentum $\hbar\mathbf{k}$ on which external forces \mathbf{F}_{ext} act. However, $\hbar\mathbf{k}$ is not the electron momentum and the right-hand side of those equations does not really contain all the external forces, since it misses the one due to the periodic potential, and embodied in the energy band. In essence thus, the dynamical behavior of quantum particles in crystals in the presence of slowly-varying external fields on the scale of a cell size, can be determined by using the following procedure

Procedure

Step 1. Check that the condition (4.4) be verified, along with the absence of inter-band transitions. If not, one has to revert back to (4.1) and follow the full quantum treatment for time-dependent problems.

Step 2. Get the stationary Bloch functions $\Psi_{n,\mathbf{k}}(\mathbf{r})$ and energy band $E_n(\mathbf{k})$ as basic ingredients, as e.g. determined in Chap. 2.

Step 3. Identify average position and velocity of the quantum particle by means of (4.5) and (4.6), respectively.

Step 4. Use the semiclassical equations of motion (4.9) and (4.10) to determine the time-evolved $\mathbf{k}(t)$. These are in the form of differential equations, to be solved with appropriate initial conditions.

Step 5. Insert the solution $\mathbf{k}(t)$ in (4.6) to obtain the group velocity $\mathbf{v}_{n,g}(t)$ as a function of time.

Step 6. Integrate $\mathbf{v}_{n,g}(t)$ with the appropriate initial conditions to obtain the particle trajectory.

In conclusion, a dynamical theory is available to determine the motion of single particles, in which the self-consistent spatially rapidly-varying periodic potential analyzed in Chap. 2 and generalized in Chap. 6 is treated on a quantum-mechanical basis, while the slowly-varying external fields are treated in a classical manner.

A few remarks are in order. First, the treatment can be extended to include the possibility that more bands be involved in the motion, by means of suited complications [5]. Equations (4.6) and (4.10) embody the interaction among electrons, or more in general charge carriers, and ions that are fixed in their lattice equilibrium positions. The latter interaction is described by the periodic lattice potential, and therefore leaves the wavevector \mathbf{k} unchanged. Additional terms could be necessary to interpret the experimental data: especially significant are found to be the electron-phonon and electron-impurities scattering terms, while electron-electron scattering often leads to immaterial effects.

Examples

In the following, the above procedure is used to calculate electron trajectories in selected cases, where the inclusion of interaction terms becomes manifest.

4.2.1 Motion in a static electric field

In this case, expression (4.10) becomes

$$\hbar \frac{d\mathbf{k}}{dt} = -e\mathbf{E}. \quad (4.11)$$

After integrating with initial condition $\mathbf{k}(0) = \mathbf{k}_0$, one obtains

$$\mathbf{k}(t) = \mathbf{k}_0 - \frac{e\mathbf{E}}{\hbar}t, \quad (4.12)$$

where it is understood that the time evolution of the quasi momentum is considered modulus the Brillouin zone. That is, once the BZ boundary is reached, the wavepacket undergoes Bragg scattering while \mathbf{k} is translated by a reciprocal lattice vector \mathbf{G} . Expression (4.12) has now to be inserted into (4.6). But $E_n(\mathbf{k})$ is a periodic function in \mathbf{k} , and periodic is also its derivative $\nabla E_n(\mathbf{k})$. This is to the striking and useful phenomenon of Bloch oscillations.

4.2.1.1 Bloch oscillations

Consider the simple one-dimensional case first. The problem can be easily solved: the linear time-dependent $k(t)$ makes the group velocity (4.6) a periodic function of time, with period amounting to $T_B = 2\pi\hbar/aF$ and frequency $\omega_B = aF/\hbar$ in terms of the constant force $F = eE$. In essence,

Concept

Bloch oscillations occur when a wavepacket in periodic potentials is subjected to the action of a constant external force. The wavepacket executes periodic oscillations within the BZ: it spans the whole BZ and each time the wavevector $k(t)$ reaches a BZ boundary, it is Bragg reflected back. The period T_B is determined by the strength of the external force and by the lattice spacing a .

Thus, the solution $k(t)$ turns out to be in the form of a sawtooth function, in particular a series of straight segments periodically repeating in time with period T_B , the values of $k(t)$ running within each period in the interval $[-\pi/a, \pi/a]$. A schematic illustration of this behavior is provided in Fig. 4.1 in the case of negligible effects due to scattering phenomena, to be introduced later on. Two remarks are interesting. First, notice that the changes of electron wavevector within any arbitrary time interval are the same for each electron wavepacket, implying that no two electrons with the same spin may occupy the same Bloch state. Second, the oscillations are both in reciprocal and in real space. Indeed, the periodicity of the band energy $E_n(\mathbf{k})$ with respect to \mathbf{k} implies that the wave-packet group velocity is periodic with respect to \mathbf{k} as well: indeed, the group velocity evolves as $\mathbf{v}_{ng}(t) = \hbar^{-1} \nabla E_n(\mathbf{k})|_{\mathbf{k}=\mathbf{k}(t)}$. Reverting back to the simple one dimensional case, we thus get that the electron position changes according to:

$$r(t) - r(0) = \int_0^t dt' v_{ng}(t') = \int_0^t \frac{dE_n(k)}{\hbar dk/dt'} = \frac{E_n(k(t)) - E_n(k(0))}{-eE}, \quad (4.13)$$

after use of (4.11). Therefore, the average semiclassical position performs periodic oscillations.

Bloch oscillations can nowadays be experimentally observed in systems that allow on-purpose design and precise control of governing parameters. Fig. 4.2 illustrates for example the first observation of Bloch oscillations of ultracold atom gases

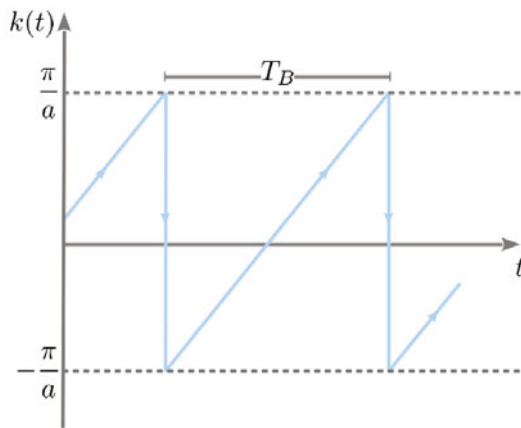
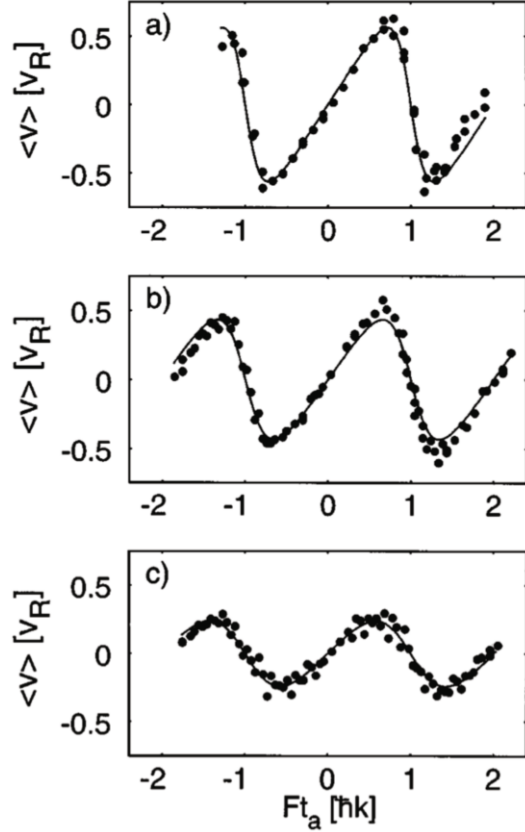


Fig. 4.1 Motion under the action of a constant external electric field. The variation of the electron wave number as a function of time, in the one-dimensional case. Scattering phenomena are neglected. The oscillation period and the Brillouin zone boundaries are evidenced.

Fig. 4.2 Transport behavior under the action of a constant force. Bloch oscillations of ultracold Cs atomic gases in periodic potentials provided by stationary waves of counter-propagating laser light. The spacing of this optical lattice is $a = \lambda_L/2$ in terms of the laser wavelength λ_L . Each dot represents one measurement of average atomic velocity $\langle v \rangle$ performed at time t_a during which the atomic sample is accelerated by a constant force F . $\langle v \rangle$ is in units of the recoil velocity $v_R = \hbar k/m$, with $k = 2\pi/\lambda_L = \pi/a$. Ft_a is in units of $\hbar k = \hbar\pi/a$, so that $F/(\hbar k) = 2/T_B$. Different panels refer to different values of the lattice potential depth U_0 in units of the recoil energy $E_R = \hbar^2 k^2/(2m)$ with m the atom mass: U_0 is indeed tunable after changes of the laser-light intensity. From (a) to (c): $U_0=1.4, 2.3$, and $4.4 E_R$ [6]. Reprinted with permission from [6]. Copyright (1996) by the American Physical Society



in periodic potentials, that are provided by stationary waves of counter-propagating laser light. The measured average atomic velocity is shown as a function of time during which the atomic wavepacket is accelerated by appropriate motion of the retro-reflecting mirrors that make the stationary laser light.

Bloch oscillations can also be viewed as the consequence of translational symmetry breaking originated by the constant force F and the correspondingly linear potential in the position: Bloch functions no longer represent eigenstates of the Hamiltonian and the energy spectrum is actually a ladder of energies related to the Bloch band energy by $\tilde{E}_{n,\ell}(\mathbf{k}) = E_n(\mathbf{k}) + \ell a F = E_n(\mathbf{k}) + \ell \hbar \omega_B$, with ℓ an integer number. The corresponding ℓ states of the ladder have a resonant character and are known as Wannier-Stark states. The energy difference between adjacent Wannier-Stark states is precisely the Bloch energy $\hbar \omega_B$, as depicted in Fig. 4.3.

Observation of Bloch oscillations. This particular view implies that observation of Bloch oscillations needs a few basic conditions. First, the initial state has to be a coherent superposition of at least two of these Wannier-Stark states, otherwise the dynamics would reduce to a trivial change in time of the global phase of the whole

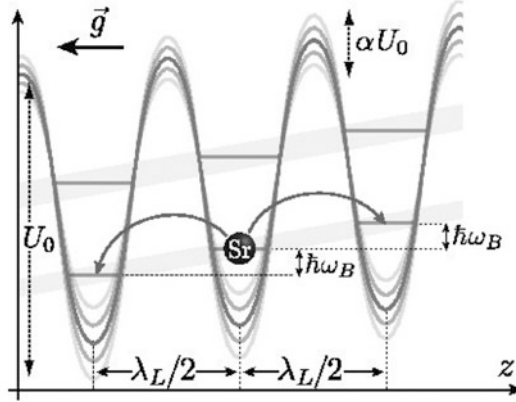


Fig. 4.3 Transport behavior under the action of a constant force. Bloch oscillations. Schematic representation of the Wannier-Stark ladder for a periodic potential with lattice spacing a tilted because of an external constant force F : adjacent Wannier-Stark states are separated by $\hbar\omega_B$ in energy. The scheme refers to the experimental setting for the measurement of Bloch oscillations with ultracold Sr atoms cooled at about 600 nK and surfing onto a vertical lattice of stationary laser light, detuned from Sr atomic transitions, under the action of the force of gravity [8]. The wiggling curves represent a modulation of the optical lattice amplitude by multiples m of the Bloch frequency ω_B , as described in the main text [8]

wave function. In this view, it is evident how the manifestation of significant Bloch oscillations requires site-to-site coherence and thus that the wavepacket remains coherent on a length scale larger than the lattice spacing. In fact, (4.11) and (4.12) do not contain dissipative terms.

Second, once averaged over the BZ, the velocity is zero, since it is an odd function of \mathbf{k} , thus posing a basic condition to observe Bloch oscillations. Two limiting cases might occur. Usually, T_B is much longer than the typical scattering time due to interactions with phonons, impurities, or other mechanisms as those discussed in Sec. 4.2.6, so that in this case wavepacket motion is interaction-driven and Bloch oscillations are not visible. If T_B is instead much shorter than typical scattering times, then the wavepacket is able to run the whole BZ and Bloch oscillations can be possibly observed. This is the case of appropriately designed nanostructure superlattices [7] and of ultracold atoms in optical lattices originated by the interference pattern of two counter-propagating laser beams with wavelength $\lambda_L = 2\pi/k_L$ in terms of the laser wavevector k_L , the lattice spacing resulting to be $a = \lambda_L/2$ [6, 8]. In both systems, scattering mechanisms are either overcome by designed-induced enhanced coherence, as in nanostructure superlattices, or are essentially absent, as in ultracold atoms. In more recent experiments with ultracold atoms, the combination of absence of scattering mechanisms in optical lattices and of dynamically enhanced coherence of the initial wave packet have led to the observation of Bloch oscillations lasting up to 17 seconds, as spectacularly displayed in Fig. 4.4 [10] with neater cleanliness with respect to those displayed in Fig. 4.2. Notice that the Bloch period is of the order of 2 ms , so that almost 10000 oscillations can be followed, reproducing at a quan-

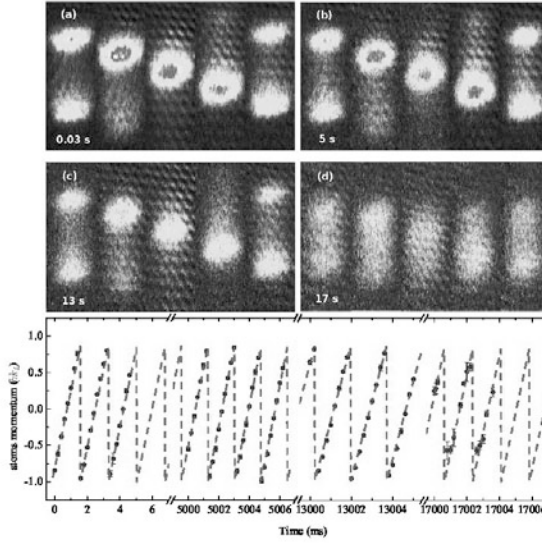


Fig. 4.4 Transport behavior under the action of a constant force. Observed Bloch oscillations in the experimental scheme of Fig. 4.3, where ultracold Sr atoms are cooled down at about 600 nK into a vertical lattice of stationary laser light, detuned from Sr atomic transitions, under the action of the force of gravity [8]. The amplitude of the optical lattice is modulated at multiples m of the Bloch frequency ω_B : the availability of m energy quanta $\hbar\omega_B$ enhances the coherent delocalization of the atomic wavepacket over m lattice sites, so that extraordinarily long-living Bloch oscillations are observed, up to 17 seconds. These are in spectacularly visible, with neater cleanliness with respect to those displayed in Fig. 4.2. Top panel: snapshot images of the atomic cloud momentum distribution obtained after absorption of light at the selected times 0.03, 5, 13, and 17 s from (a) to (d) [9]. Bottom panel: dots with error bars represent the measured average atom momentum as a function of time in ms [10]. Notice that the Bloch period is of the order of 2 ms , so that up to 10^4 thousands oscillations can be followed, reproducing at a quantum level the pendulum experiment with which Galileo first tested the equivalence principle between inertial and gravitational mass. This mechanism is relevant for quantum-transport applications and fundamental-physics tests [8]. Reprinted with permission from [8]. Copyright (2012) by the American Physical Society

tum level the pendulum experiment with which Galileo first tested the equivalence principle between inertial and gravitational mass. Bloch oscillations are indeed especially useful to design quantum-transport devices [8] and to perform precision measurements for fundamental physics tests [11].

Ohm's law. The 3D case offers two different limiting possibilities and leads to an important concept. In fact, once the BZ boundary is reached and a Bragg reflection occurs, the wavepacket does not necessarily enters the BZ through the same set of states: an average has therefore to be performed over a multiplicity of trajectories in the BZ. In one limiting situation, no trajectory is covered twice and therefore the curves set fills in the BZ. In the other, a reciprocal wave vector might exist, that is parallel to the electric field connecting the opposite ends of the trajectory, so that after a transient time the wavepacket starts following the same curve in the BZ. In both cases the average electron velocity vanishes along with the electric current.

Therefore, the system described by (4.11) is not consistent with Ohm's law and the existence of resistance to current flow. As it is discussed later on in this Chapter, this occurs because (4.11) misses the occurrence of scattering events. Thus

Concept

Resistance to the passage of electric current in materials is a concept connected to scattering events.

The above results can be extended to bands containing more than one electron and partially filled. In the case of completely filled bands, the instantaneous current is identically zero, even when scattering is taken into account and provided that no transitions to other bands occur. Indeed, if N electrons completely fill the states belonging to the given band, the following occurs:

Properties

P1. The total wave vector associated to the band is zero, that is $\mathbf{k}_{\text{tot}} = \sum_{i=1}^N \mathbf{k}_i = \mathbf{0}$, in terms of the wavevector \mathbf{k}_i associated to the i -th electron.

P2. The net velocity is zero, so that no net flux of current may manifest. Indeed, the group velocity $\mathbf{v}_{ng}(\mathbf{k})$ is an odd function of \mathbf{k} . If the band is completely filled, the contribution of an electron with velocity $\mathbf{v}_{ng}(\mathbf{k})$ is exactly compensated by that of an electron with velocity $\mathbf{v}_{ng}(-\mathbf{k}) = -\mathbf{v}_{ng}(\mathbf{k})$.

P3. For each electron, the time-average wavevector $T_B^{-1} \int_t^{t+T_B} dt' \mathbf{k}(t')$ is zero, as evidently represented in Fig. 4.1.

In conclusion,

Concept

Filled electronic bands do not contribute to electric currents. From this consideration the classification of solids in metals and insulators arises, according to whether electronic bands are completely or partially filled.

On the other hand, semiconductors can be viewed as a particular class of insulators, since they do have a band structure at $T = 0$ K similar to that typical of insulators, though the energy gap is sufficiently small for the electrons to be excited by thermal or optical means into the conduction band at room temperature. That is, a finite conductivity can be measured after, e.g., an increase of temperature. A

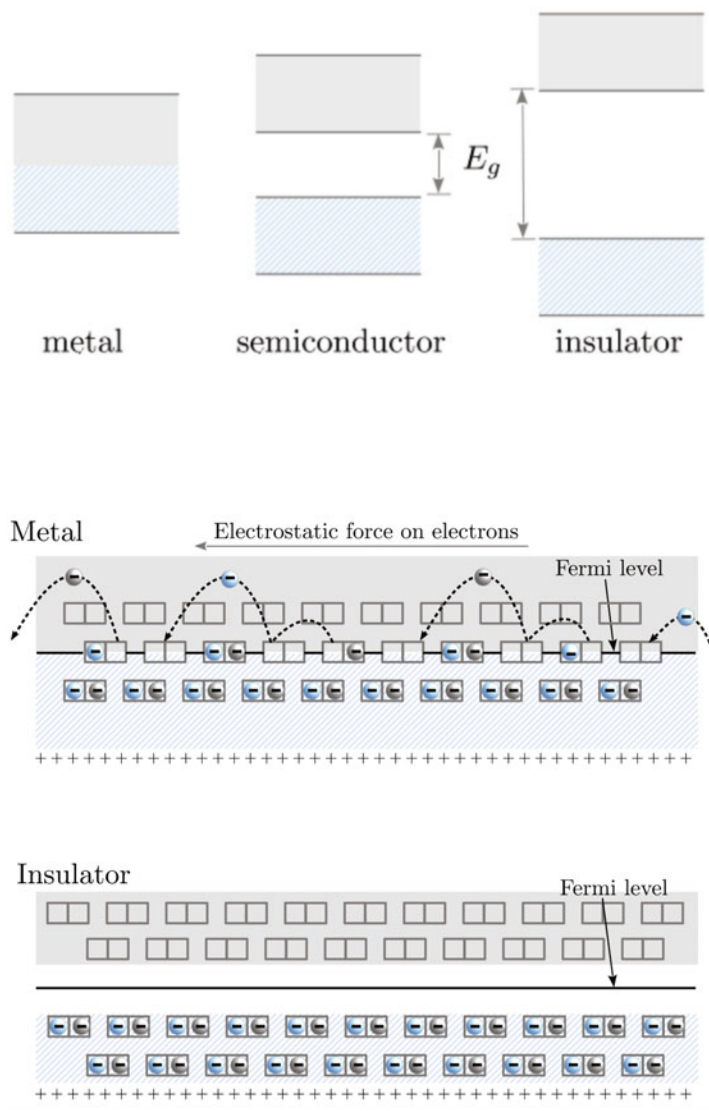


Fig. 4.5 Top: classification of solids according to their transport properties: metals, insulators and semiconductors can be defined according to the occupation of bands and the magnitude of the energy gap. The cyan hatched (grey shaded) regions correspond to filled (empty) electronic states. Bottom: a schematic view of the effect of an applied electric field [1]. In a metal, the Fermi level lies in the middle of a band. As such, electrons can shift through the metal, with the help of the empty levels. A net flow of charges takes place. On the other hand, in an insulator or a semiconductor at zero temperature, the Fermi level lies in the middle of a forbidden energy gap. No empty levels are available to the electrons of the filled bands, and no net charge flow can occur.

schematic picture is given in Fig. 4.5. One of the main goals of band theory has indeed been to explain why the resistivity of a material can span a range of about

30 orders of magnitude, from the typical values $\approx 10^{22} \Omega \cdot \text{cm}$ of insulators down to $\approx 10^{-10} \Omega \cdot \text{cm}$ of pure metals at low temperatures.

Quick Questions

Q1. A neutral superfluid system in a periodic potential is characterized by the energy-band dispersion $E_n(\mathbf{k}) = c|\mathbf{k}|$, with c a constant. Determine the $\mathbf{v}_{ng}(t)$ at the BZ center.

Answer - The derivative $\hbar^{-1}\nabla E_n(\mathbf{k})$ does not exist at $k = 0$. In this case one has to revert back to the calculation of the average of the quantum momentum operator $-i\hbar\nabla$ on the Bloch functions. The semiclassical approximation could be used at a second stage, via (4.12).

Q2. If the average time lasting between two successive scattering events

of an electron with a lattice vibration is $\tau_0 \simeq 10^{-13}$ s, assess the conditions under which Bloch oscillations are visible in a 1D crystal.

Answer - Be a the lattice parameter and E the applied electric field. The suggested time T_B is given by $\hbar \cdot 2\pi/a = eET_B$, that is, $T_B = 2\pi\hbar/(eaE)$. Visibility of Bloch oscillations requires $T_B < \tau_0$, thus $E > 2\pi\hbar/(ea\tau_0) \simeq 10^4$ (CGS units).

Q3. Are Bloch oscillations meaningful in a 3D crystal?

Answer. Yes, they do whenever a regime of periodicity is somehow established for the time-evolution of $\mathbf{k}(t)$ in the BZ.

Q4. Bloch oscillations in real space have been observed in systems of ultracold atoms moving in optical lattices. Discuss the least conditions that should have been engineered on the energy bandwidth to make them visible.

Answer. Oscillations of $\mathbf{v}_{ng}(t) = \hbar^{-1}\nabla E_n(\mathbf{k})|_{\mathbf{k}=\mathbf{k}(t)}$ correspond to oscillations of $\tilde{\mathbf{R}} = \int dt' \mathbf{v}_{ng}(t')$. Therefore, elaborating on expression (4.13) shows that visibility of Bloch oscillations in real space are critically dependent on the fact that the bandwidth should be large on the scale of the Bloch energy Fa .

Q5. Since each band is composed by N states corresponding to the total number N of elementary cells in the selected volume, one could conclude that a material results to be insulator if each atom in the crystal has an even number of valence electrons. Otherwise, the material turns out to be a metal. Discuss whether this is necessarily so.

Answer. It is not necessarily so in 3D because of the very common characteristics of three-dimensional crystal bands to show degeneracy, in contrast with one-dimensional crystals where degeneracy is not a possibility.

4.2.2 Motion in a static magnetic field

Consider now the case in which only a static magnetic field \mathbf{B} , say $\mathbf{B} = B\hat{\mathbf{B}}$, is present, with $\hat{\mathbf{B}}$ versor of \mathbf{B} . Thus expression (4.10) becomes

$$\hbar \frac{d\mathbf{k}}{dt} = -\frac{e}{\hbar c} \nabla E_n(\mathbf{k}) \wedge \mathbf{B}, \quad (4.14)$$

along with

$$\hbar \frac{d\mathbf{r}}{dt} = \frac{1}{\hbar} \nabla E_n(\mathbf{k}), \quad (4.15)$$

These laws are especially relevant, since they may provide information on the electronic bands $E_n(\mathbf{k})$. In the following this possibility is explored. In fact, (4.14) implies that variations $d\mathbf{k}$ of \mathbf{k} in an infinitesimal time dt are perpendicular to both group velocity and magnetic field. Be E the initial electron energy. From geometry and differential calculus one gets that for any given energy E , the expression $E_n(\mathbf{k}) = E$ defines a surface in the space of \mathbf{k} vectors with the following characteristics: vector $\nabla E_n(\mathbf{k})$ is perpendicular to the surface in whatever point of the surface itself. Thus, the solution $\mathbf{k}(t)$ of Eq. (4.14) is obtained after intersecting the constant-energy surface defined above with a plane perpendicular to the given magnetic field direction. This is schematically shown in Fig.4.6. The position of the plane perpendicular to \mathbf{B} is determined by the condition that the component k_{\parallel} of the \mathbf{k} vector along \mathbf{B} is to be constant. In the case of free particles, this latter condition would imply that the velocity be constant along $\hat{\mathbf{B}}$. In the actual case though, (4.6) shows that $\mathbf{v}_{n,g}$ is a function of all the \mathbf{k} components. The function $\mathbf{k}(t)$ that solves (4.14), can result to be a either closed or open trajectory, depending on the shape of the constant-energy surface and on the intersecting plane. As an example, situations are possible in which different directions of the plane perpendicular to \mathbf{B} could lead to open or closed trajectories, once the constant-energy surface has been fixed. In

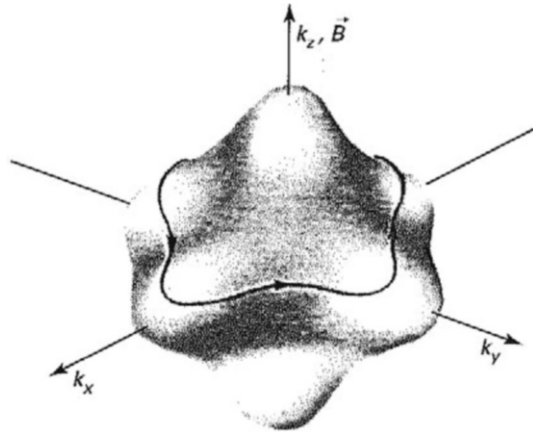


Fig. 4.6 Motion of an electron in a static magnetic field. The solution $\mathbf{k}(t)$ of (4.14) is obtained after intersecting the constant-energy surface defined above with a plane perpendicular to the given magnetic field direction. The magnetic field is directed along the k_z

the closed case, the trajectory would be traveled clockwise or counter-clockwise, depending on whether $\nabla E_n(\mathbf{k})$ be directed outward or inward with respect to the surface. In summary, the solution of (4.14) thus requires the knowledge of $E_n(\mathbf{k})$. The trajectory $\mathbf{k}(t)$ is connected to the projection $\mathbf{r}_\perp(t)$ of the electron trajectory $\mathbf{r}(t)$ onto a plane perpendicular to \mathbf{B} . One has:

$$\begin{aligned}\hat{\mathbf{B}} \wedge \hbar \frac{d\mathbf{k}}{dt} &= -\frac{eB}{c} (\dot{\mathbf{r}}(t) - \hat{\mathbf{B}}(\hat{\mathbf{B}} \cdot \dot{\mathbf{r}}(t))) = -\frac{eB}{c} \dot{\mathbf{r}}_\perp(t), \\ \mathbf{r}_\perp(t) &= \mathbf{r}(t) - \hat{\mathbf{B}}(\hat{\mathbf{B}} \cdot \mathbf{r}(t)).\end{aligned}\quad (4.16)$$

After integrating, one obtains

$$\mathbf{r}_\perp(t) - \mathbf{r}_\perp(0) = -\frac{\hbar c}{eB} \hat{\mathbf{B}} \wedge (\mathbf{k}(t) - \mathbf{k}(0)), \quad (4.17)$$

showing that

Concept

The projection $\mathbf{r}_\perp(t)$ on a plane perpendicular to \mathbf{B} of an electron trajectory in real space and the trajectory $\mathbf{k}(t)$ in k -space are reciprocally connected. This concept turns out to be very useful in characterizing the band structures under suited conditions. The trajectories are traveled with angular frequency $\omega_c = eB/(m^*c)$. The quantities ω_c and m^* are named cyclotron frequency and effective cyclotron mass, respectively. This can be exploited to infer information on the surfaces at constant energies along different BZ directions.

The question arises on how to characterize the trajectories, so that information on the material band structure can be extracted. The simplest example is now given, that is afterward coded into a procedure.

Examples

In the case of free particles, the surfaces with constant energy are defined by the expression $E = E(k) = \frac{\hbar^2}{2m} (k_\perp^2 + k_\parallel^2)$, where k_\perp and k_\parallel are the components of \mathbf{k} onto the perpendicular and parallel to \mathbf{B} planes, respectively. These surfaces are clearly spheres. The intersection with the planes perpendicular to \mathbf{B} thus originates closed circular trajectories with radius $k_\perp = \sqrt{2mE/\hbar^2 - k_\parallel^2}$. Notice that ω_c is the same for all the trajectories.

The question is now faced to determine the cyclotron frequency of closed orbits in the case of more complex surfaces. To this aim, one first grabs from (4.17) the concept that the period of the real-space trajectories $\mathbf{r}_\perp(t)$ is the same as that of the k -space trajectories $\mathbf{k}(t)$. Then, the following procedure can be followed, that is valid within the semiclassical approximation:

Procedure

Step 1. Calculate the time $t_2 - t_1$ needed to travel a segment of trajectory $\mathbf{k}(t)$. This is:

$$t_2 - t_1 = \int_{\mathbf{k}_1}^{\mathbf{k}_2} \frac{dk}{|\dot{\mathbf{k}}|} = \frac{\hbar^2 c}{eB} \int_{\mathbf{k}_1}^{\mathbf{k}_2} \frac{dk}{|\nabla E_n(\mathbf{k})_{\perp}|}, \quad (4.18)$$

with $\nabla E_n(k)_{\perp}$ the projection of $\nabla E_n(\mathbf{k})$ onto the plane where the trajectory $\mathbf{k}(t)$ lies.

Step 2. Manage the integral (4.18) by relating the energy distance between two close trajectories at E and $E + \Delta E$ in terms of their \mathbf{k} -space distance $\Delta \mathbf{k}$. This is:

$$\Delta E = \nabla E_n(\mathbf{k})_{\perp} \cdot \Delta \mathbf{k} = |\nabla E_n(\mathbf{k})_{\perp}| \Delta k, \quad (4.19)$$

since $\nabla E_n(k)_{\perp}$ and Δk are parallel to each other. Thus

$$t_2 - t_1 = \frac{\hbar^2 c}{eB} \int_{\mathbf{k}_1}^{\mathbf{k}_2} \frac{dk \Delta k}{\Delta E} = \frac{\hbar^2 c}{eB} \frac{1}{\Delta E} \int_{\mathbf{k}_1}^{\mathbf{k}_2} dk \Delta k. \quad (4.20)$$

The integral on the right-hand side of (4.20) represents the area A of that portion of plane perpendicular to \mathbf{B} , that is enclosed by the two trajectories with energy difference ΔE .

Step 3. Perform the $\Delta E \rightarrow 0$ limit of (4.20) to calculate the traveling time related to the given trajectory. This is:

$$T = \frac{\hbar^2 c}{eB} \frac{\partial A}{\partial E}. \quad (4.21)$$

Step 4. Cast (4.21) in the form $T = 2\pi m^* c / (eB)$ and define the effective cyclotron mass

$$m^* = \frac{1}{2\pi} \hbar^2 \frac{\partial A}{\partial E}. \quad (4.22)$$

In general, in the vicinity of critical points m^* differs from the effective band mass. The free-particle example is consistent with the above equations: in particular, one has $A = \pi k_{\perp}^2 = \pi (2mE/\hbar^2 - k_{\parallel}^2)$, thus $\partial A / \partial E = 2\pi m / \hbar^2$ and (4.21) yields the cyclotron period $T = 2\pi / \omega_c$, as already found. At variance with the case of free particles though, in (4.21) the traveling time is not the same for all the trajectories: this result is exploited in experiments, to extract information on the surfaces with constant energies along different directions of the BZ.

4.2.2.1 Cyclotron and band effective mass

One may ask whether cyclotron and band effective mass bear similar or different information on band structures. Consider a minimum or a maximum of the electron band. In the vicinity of either one of the two points, say \mathbf{k}_0 , one can expand the energy to second order in wavevector \mathbf{k} as:

$$E_n(\mathbf{k}) = E_n(\mathbf{k}_0) + \hbar^2(\mathbf{k} - \mathbf{k}_0) \cdot \mathbf{M}^{-1} \cdot (\mathbf{k} - \mathbf{k}_0), \quad (4.23)$$

where \mathbf{M}^{-1} represents the inverse mass tensor. Operating by means of a rotation (orthogonal) transformation of the reference system, (4.23) becomes

$$E_n(\mathbf{k}) = E_n(\mathbf{k}_0) + \hbar^2 \left(\frac{k_x^2}{2m_1} + \frac{k_y^2}{2m_2} + \frac{k_z^2}{2m_3} \right), \quad (4.24)$$

with m_1, m_2 and m_3 the principal effective masses, that are either positive or negative quantities depending on whether \mathbf{k}_0 is a minimum or a maximum point for the band. Now, the surfaces with constant energy are ellipsoids. In addition, once the direction of \mathbf{B} is fixed, say along $\hat{\mathbf{z}}$, k_z is constant and the intersection with a plane orthogonal to \mathbf{B} determines the elliptic trajectory

$$\hbar^2 \left(\frac{k_x^2}{2m_1} + \frac{k_y^2}{2m_2} \right) = E', \quad (4.25)$$

with $E' = E - \hbar^2 k_z^2 / (2m_3) - E_n(k_0)$ for the point of band minimum and $E' = E_n(k_0) - \hbar^2 k_z^2 / (2|m_3|) - E$ for the point of band maximum. Knowledge of the principal ellipse axes $a^2 = 2|m_1|E'/\hbar^2$ e $b^2 = 2|m_2|E'/\hbar^2$ determines the area $A = \pi ab = 2\pi E' \hbar^{-2} \sqrt{|m_1 m_2|}$ and its energy derivative $\partial A / \partial E = 2\pi \hbar^{-2} \sqrt{|m_1 m_2|}$. The effective cyclotron mass is thus $m^* = \sqrt{|m_1 m_2|} = (|m_1 m_2 m_3| / |m_3|)^{1/2}$. The numerator can be cast in the form $\det |\mathbf{M}|$, making evident that it is invariant after orthogonal transformations. One may conclude that $m^* = (\det |\mathbf{M}| / |m_3|)^{1/2}$, generally different from each of the three masses $|m_1|, |m_2|$ e $|m_3|$. Thus

Concept

Cyclotron and band effective masses are in general different in the vicinity of critical points. The difference is in essence originated by the fact that the cyclotron motion involves two among the three components of \mathbf{k} and thus the corresponding inertia results from a geometrical average of effective masses. The central result is that traveling periods of different orbits are different, providing an useful tool to determine the shape of Fermi surfaces, especially in metals.

4.2.2.2 Particles with spin

First, it is recalled from Chap. 2 the Curie law $M_{\parallel} = Ng^2\mu_b^2 j(j+1)/(3k_b T)$, stating how the magnetization scales inversely with temperature T . Two next spin states in the manifold, say with number of atoms per unit volume N_+ at higher and N_- at lower energy, populate according to the ratio of their energy difference $\hbar\omega_0$ to the thermal energy $k_b T \equiv 1/\beta$, that is $N_+/N_- = e^{-\beta\hbar\omega_0}$. The total equilibrium magnetization of the sample is therefore $M_0 = (N_- - N_+)\mu = N\mu \tanh(\beta\hbar\omega_0/2)$.

Second, the angular momentum evolves in time driven by the torque $\boldsymbol{\mu} \wedge \mathbf{B}$, according to $\hbar d\mathbf{I}/dt = \boldsymbol{\mu} \wedge \mathbf{B}$. In terms of the magnetization, this is

$$\frac{d\mathbf{M}}{dt} = -\frac{g\mu_b}{\hbar} \mathbf{M} \wedge \mathbf{B}. \quad (4.26)$$

Equation (4.26) is clearly incomplete, as it predicts that a sample placed in a magnetic field would instantaneously adjust to build up its magnetization from zero to the equilibrium value M_0 after the field is switched on. This cannot really be the case, in fact, since characteristic relaxation times are to be introduced. This is the topic of Sec. 4.2.6.2.

4.2.3 Motion in the presence of orthogonal electric and magnetic fields

Consider now the complete equation

$$\hbar \frac{d\mathbf{k}}{dt} = -\frac{e}{\hbar c} \nabla E_n(\mathbf{k}) \wedge \mathbf{B} - e\mathbf{E}, \quad (4.27)$$

representing the case where both electric and magnetic fields act. Using the same derivation as that leading to (4.16), one has:

$$\mathbf{r}_{\perp}(t) - \mathbf{r}_{\perp}(0) = -\frac{\hbar c}{eB} \hat{\mathbf{B}} \wedge [\mathbf{k}(t) - \mathbf{k}(0)] + \mathbf{w}t, \quad (4.28)$$

where $\mathbf{w} = cE\hat{\mathbf{E}} \wedge \hat{\mathbf{B}}/B$. The trajectory $\mathbf{r}_{\perp}(t)$ is once again related to the trajectory $\mathbf{k}(t)$, though the additional term $\mathbf{w}t$ is now appearing, which represents an uniform motion of the electron in the direction perpendicular to both \mathbf{B} and \mathbf{E} .

Implications of (4.28) are especially relevant when charge carriers travel closed orbits with period T long on the scale of the inverse cyclotron frequency, that is $\omega_c T \gg 1$ or else large magnetic field, so that the carrier may wind around many times. For time intervals $\tau > T$ one has

$$\frac{\mathbf{r}_{\perp}(\tau) - \mathbf{r}_{\perp}(0)}{\tau} = -\frac{\hbar}{m\omega_c} \hat{\mathbf{B}} \wedge \frac{[\mathbf{k}(\tau) - \mathbf{k}(0)]}{\tau} + \mathbf{w}. \quad (4.29)$$

In the absence of electric fields, the trajectory $\mathbf{k}(t)$ is closed. If in addition the trajectory is negligibly shifted during time τ , the term on the left-hand side of (4.29) approximately represents the velocity of the electron in the plane perpendicular to the magnetic field, and the first term on the right-hand side can be neglected. The current density in that plane therefore turns out to be

$$\mathbf{j}_{\perp} = -ne\mathbf{w} = -\frac{ne\hbar}{B}\mathbf{E} \wedge \hat{\mathbf{B}}. \quad (4.30)$$

The current \mathbf{j}_{\perp} results to be orthogonal to both the electric and magnetic fields, the component of \mathbf{E} along the magnetic field playing no significant role. This result turns out to be useful later on in the section devoted to metals, where the motion of free particles is investigated in the presence of viscous forces arising from scattering mechanisms and under particular geometrical conditions, leading to the Hall effect.

4.2.4 A generalized Newton law: the effective mass tensor

The concept of effective mass tensor is more generally useful in understanding the electron dynamics. To this aim, the definition is now extended to any band point, beyond the critical ones for which it has been introduced so far. Yet within the validity of semiclassical approximation, expression (4.6) can be used to determine the particle acceleration as:

$$\mathbf{a}_{n,g} = \frac{1}{\hbar} \frac{d}{dt} \nabla E_n(\mathbf{k}). \quad (4.31)$$

The i -th component of the acceleration is

$$\begin{aligned} a_{n,g,i} &= \frac{1}{\hbar} \frac{d}{dt} \frac{\partial}{\partial k_i} E_n(\mathbf{k}) = \frac{1}{\hbar} \sum_j \frac{\partial^2}{\partial k_j \partial k_i} E_n(\mathbf{k}) \dot{k}_j \\ &= \frac{1}{\hbar^2} \sum_j \frac{\partial^2}{\partial k_j \partial k_i} E_n(\mathbf{k}) F_{ex,j}. \end{aligned} \quad (4.32)$$

Definition

The tensor \mathbf{M}^{-1} with components

$$M_{i,j}^{-1} = \frac{1}{\hbar^2} \frac{\partial^2 E_n(\mathbf{k})}{\partial k_j \partial k_i}, \quad (4.33)$$

is the inverse mass tensor, the components having the dimensions of an inverse mass.

The inverse tensor \mathbf{M} satisfies the eloquent equation

$$\mathbf{M} \cdot \mathbf{a} = \mathbf{F}_{\text{ext}}, \quad (4.34)$$

defining the role of the effective mass tensor: it embodies the periodic potential contributions and determines the particle acceleration under the action of external forces.

Expression (4.34) is here used to calculate the cyclotron mass in the presence of a magnetic field with arbitrary direction, according to the following

Procedure

Step 1. Consider the vector relation

$$\mathbf{M} \cdot \frac{d\mathbf{v}_{n,g}}{dt} = -\frac{e}{c} \mathbf{v}_{n,g} \wedge \mathbf{B}, \quad (4.35)$$

and write it in terms of its components with respect to the principal axes of the \mathbf{M} tensor. Omitting indexes n and g for simplicity, one has

$$\begin{aligned} m_1 \frac{dv_x}{dt} &= -\frac{e}{c} (\mathbf{v} \wedge \mathbf{B})_x \\ m_2 \frac{dv_y}{dt} &= -\frac{e}{c} (\mathbf{v} \wedge \mathbf{B})_y \\ m_3 \frac{dv_z}{dt} &= -\frac{e}{c} (\mathbf{v} \wedge \mathbf{B})_z. \end{aligned} \quad (4.36)$$

Step 2. Use the ansatz $\mathbf{v}(t) = \mathbf{u}e^{i\omega t}$, and turn the above system of differential equations into a homogeneous linear system of algebraic equations for the unknown \mathbf{u} components.

Step 3. In calculating the determinant of the coefficients, use the effective cyclotron mass $\omega_c = eB/(m^*c)$ and express $\mathbf{B} = (B \sin \theta \cos \phi, B \sin \theta \sin \phi, B \cos \theta)$ in polar coordinates with the z -axis along the m_3 direction. As shown in Problem 4.1, it results that the determinant of the coefficients vanishes if

$$m^{*-1} = \left(\frac{\sin^2(\theta) \cos^2(\phi)}{m_2 m_3} + \frac{\sin^2(\theta) \sin^2(\phi)}{m_1 m_3} + \frac{\cos^2(\theta)}{m_1 m_2} \right)^{\frac{1}{2}}. \quad (4.37)$$

Step 4. Given $E_n(\mathbf{k})$, compute the acceleration (4.31) and relate it to the effective mass (4.37) or viceversa.

The result (4.37) generalizes the expression already found in the case where the field was directed along the z -axis. In the following, a selected limiting cases are discussed. In the case with $m_1 = m_2$, (4.37) leads to $m^{*-1} = m_1^{-\frac{1}{2}} (m_3^{-1} \sin^2 \theta + m_1^{-1} \cos^2 \theta)^{\frac{1}{2}}$. The cyclotron effective mass has cylindrical symmetry around the z -axis. In the case with $m_1 = m_2 = m_3$, one gets the isotropic mass $m^* = m_1 = m_2 = m_3$.

Consider now the following applications of the procedure above.

Examples

Consider the case of the simple-cubic lattice, that is with the typical band structure

$$E(\mathbf{k}) = E_0 (\cos(k_x a) + \cos(k_y a) + \cos(k_z a)).$$

It can be easily shown that $M_{i,j}^{-1} = 0$ for $i \neq j$, whereas

$$M_{11}^{-1} = -E_0 a^2 \hbar^{-2} \cos(k_x a) \Rightarrow M_{11} = \hbar^2 [-E_0 a^2 \cos(k_x a)]^{-1},$$

$$M_{22}^{-1} = -E_0 a^2 \hbar^{-2} \cos(k_y a) \Rightarrow M_{22} = \hbar^2 [-E_0 a^2 \cos(k_y a)]^{-1},$$

and

$$M_{33}^{-1} = -E_0 a^2 \hbar^{-2} \cos(k_z a) \Rightarrow M_{33} = \hbar^2 [-E_0 a^2 \cos(k_z a)]^{-1}.$$

The sign of each of the above masses is opposite to that of E_0 in the vicinity of $k = 0$, it is the same as that of E_0 in the vicinity of $(\pi/a, \pi/a, \pi/a)$. The inverse mass components become infinitely large while $k_x a = k_y a = k_z a \rightarrow \pi/2$. In addition, when k_x is close to the origin while k_y and k_z close to π/a , the sign of M_{11} is opposite to that of E_0 , whereas the sign of M_{22} and M_{33} is the same as for E_0 .

Examples

Consider now a face-centered cubic lattice, that is with the typical band structure

$$E(\mathbf{k}) = E_0 [\cos(k_x a) \cos(k_y a) + \cos(k_x a) \cos(k_z a) + \cos(k_y a) \cos(k_z a)].$$

In this case, one finds the following results:

$$M_{11}^{-1} = -E_0 a^2 \hbar^{-2} \cos(k_x a) [\cos(k_y a) + \cos(k_z a)],$$

$$M_{22}^{-1} = -E_0 a^2 \hbar^{-2} \cos(k_y a) [\cos(k_x a) + \cos(k_z a)],$$

$$M_{33}^{-1} = -E_0 a^2 \hbar^{-2} \cos(k_z a) [\cos(k_y a) + \cos(k_x a)],$$

$$M_{12}^{-1} = M_{21}^{-1} = E_0 a^2 \hbar^{-2} \sin(k_x a) \sin(k_y a),$$

$$M_{13}^{-1} = M_{31}^{-1} = E_0 a^2 \hbar^{-2} \sin(k_x a) \sin(k_z a),$$

and finally

$$M_{23}^{-1} = M_{32}^{-1} = E_0 a^2 \hbar^{-2} \sin(k_z a) \sin(k_y a).$$

The quantities $M_{i,j}$ are found after inverting the above symmetric matrix M_{ij}^{-1} .

Quick Questions

Q6. Consider the above examples. Is in any of those the possibility of finding open orbits?

Answer - Yes, there is. Each band has saddle points between the critical points. In the vicinity of the saddle points, the surfaces with constant energy are hyperboloids, whose intersection with the planes perpendicular to \mathbf{B} are in fact hyperbolas, so that the corresponding trajectories are open orbits.

4.2.5 Holes vs. electrons

Physical quantities such as the electric current are built up after averaging over the contribution of all the electrons in a given band. For example, in Sec. 4.2.1.1 it has been argued that the contribution to the total current coming from a completely filled band is zero, because after summing up the velocities of all electrons belonging to that band, no net velocity shows up. A similar procedure applies to the conduction band of a metal, discussed in Sec. 4.3, or of a semiconductor in the presence of thermally excited electrons, discussed in Sec. 4.4. On the other hand, when a band is almost filled, one could ask whether it is possible to infer the behavior of all the electron physical quantities of interest in that band by following instead the dynamics of the (few) empty states, rather than by summing over the (many) filled states. In other words, the question is whether the average procedure could be performed over fewer fictitious particles with definite mass, wavenumber and energy representing empty states, instead than over all the electrons. If these two approaches are equivalent, the results are the same with the advantage of including a fewer particles in the calculation, and the trick is worth to be used. This is discussed in the following.

Consider an otherwise filled band but with only one empty electron state having wavenumber \mathbf{k}_e and energy $E_n(\mathbf{k}_e)$. To this overall state, the name hole is given. As for an electron, this fictitious particle is identified too by its wavevector \mathbf{k}_h , mass m_h^* , and charge q_h . All these properties are determined so that the response of the hole to an external perturbation, such as e.g. an electric field, by definition works in precisely the same manner as that of the whole set of electrons belonging to that band. In particular, holes have the following properties:

Properties

P1. Wavevector. The hole wave vector is found to be

$$\mathbf{k}_h = \sum_{i \neq e} \mathbf{k}_i = \sum_{i \neq e} \mathbf{k}_i + \mathbf{k}_e - \mathbf{k}_e = -\mathbf{k}_e,$$

since the sum over all the occupied state vanishes. Thus, the hole and the missing electron wavevectors are equal in magnitude and opposite in sign:

$$\mathbf{k}_h = -\mathbf{k}_e. \quad (4.38)$$

P2. Energy. The hole energy turns out to be

$$E_n^h(\mathbf{k}_h) = \sum_{i \neq e} E_n(\mathbf{k}_i) - E_0 = \sum_{i \neq e} E_n(\mathbf{k}_i) + E_n(\mathbf{k}_e) - E_n(\mathbf{k}_e) - E_0 = -E_n(\mathbf{k}_e),$$

with $E_0 = \sum_i E_n(\mathbf{k}_i)$. Therefore, a hole in the bottom (top) of a band has a larger (lower) energy than the corresponding electron case:

$$E_n^h(\mathbf{k}_h) = -E_n(\mathbf{k}_e). \quad (4.39)$$

The contradiction is only apparent: to create a hole in the bottom of the band one has to remove the corresponding electron, and this operation requires a larger energy than creating it at the top of the same band.

P3. Velocity. Since $\nabla_{\mathbf{k}_h} E_n^h(\mathbf{k}_h) = \nabla_{\mathbf{k}_e} E_n(\mathbf{k}_e)$, the hole has the same velocity as the missing electron does.

P4. Mass tensor. As to the inverse mass tensor, one has

$$(\mathbf{M}_h^{-1})_{i,j} = \frac{1}{\hbar^2} \frac{\partial^2 E_n^h(\mathbf{k}_h)}{\partial k_{h,i} \partial k_{h,j}} = -(\mathbf{M}_e^{-1})_{i,j}.$$

Therefore, the hole mass has the same magnitude as that of the missing electron, but with opposite sign:

$$(\mathbf{M}_h^{-1})_{i,j} = -(\mathbf{M}_e^{-1})_{i,j}. \quad (4.40)$$

P5. Charge and electric current. Finally, the equation of motion for a hole is

$$\hbar \frac{d\mathbf{k}_h}{dt} = -\hbar \frac{d\mathbf{k}_e}{dt} = -\mathbf{F}_{\text{ext}} = e(\mathbf{E} + \frac{1}{c} \nabla_{\mathbf{k}_h} E_n^h(\mathbf{k}_h) \wedge \mathbf{B}). \quad (4.41)$$

This implies that under the action of electric and magnetic fields, a hole behaves as if it were a particle with positive charge. The electric current is thus $\mathbf{j} = -\sum_{i \neq e} e\mathbf{v}_i = -\sum_{i \neq e} e\mathbf{v}_i - e\mathbf{v}_e + e\mathbf{v}_e = e\mathbf{v}_e = e\mathbf{v}_h$, that is the current due to all the electrons is equivalent to the current due to the one single hole with positive charge:

$$\mathbf{j} = e\mathbf{v}_h, \quad (4.42)$$

Concept

In conclusion, whenever an almost filled band is considered, the introduction of the concept of hole is very convenient, as a particle with wavenumber, mass, energy, and charge current effectively defined in terms of the electrons that almost fill in the band.

Whenever the transport behavior within an almost filled band is to be investigated, the following procedure can be usefully adopted:

Procedure

Step 1. Identify the holes corresponding to the almost filled band.

Step 2. Express their properties according to (4.38)-(4.42).

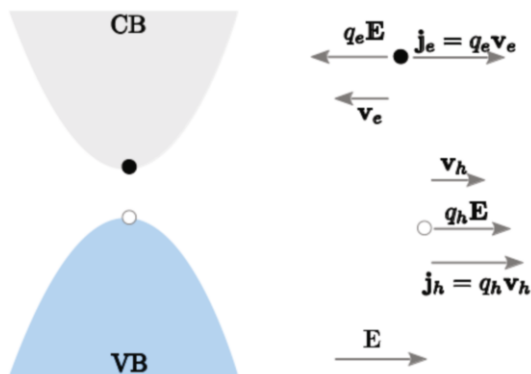
Step 3. Set the equations of motion for the holes and solve them for the equivalent problem.

Application of the above procedure is performed in the following:

Examples

Consider a semiconductor where one single electron is thermally excited from the valence to the conduction band: this is a typical process occurring in semiconductors, which is analyzed in detail in Sec. 4.4. In the presence of an applied external electric field, both bands contribute to the electric current: indeed, neither one of them is completely filled. The total current associated to the electron in the conduction band and the hole in the valence band can be easily evaluated. Electron and hole are in fact characterized by charges with opposite signs and by velocities with opposite signs too. Therefore, the electron and the hole currents share the same direction

Fig. 4.7 Transport behavior of semiconductors. Left panel: a typical process in a semiconductor. An electron (black dot) is thermally excited from the valence (VB) to the conduction (CB) band: a hole (white dot) is left in the VB. Right panel: under an applied external electric field, the CB electron and the VB hole currents have the same direction and sign



and sign, so that both carriers equally contribute to the total current. This is sketched in Fig. 4.7. It represents a fundamental and widely applied result in semiconductor physics.

Quick Questions

Q7. Tell the sign of the effective mass of an electron in the vicinity of the top of the valence band. Which sign would have a hole in the vicinity of this same top of the valence band?

Answer - The second derivative of the band dispersion implies a negative effective mass for the electron. As a consequence, the hole would have a positive effective mass instead.

Q8. Consider alternatively an electron and a hole in the vicinity of the bottom of a given valence band. What is the sign of their effective masses?

Answer. Along the same reasoning lines as in Q7, the electron would have a positive and the hole a negative effective mass.

Q9. Is a cyclotron frequency meaningful for a hole?

Answer. Yes, it is. The equation of motion for a hole (4.41) is similar to that for an electron, but the opposite sign of the charge. In the presence of magnetic fields, trajectories similar to those of electrons are thus possible, the traveling direction in \mathbf{k} -space being just opposite.

Q10. Discuss the meaning of a negative effective mass.

Answer. An electron at the top of a band has a negative mass, as it is mathematically expressed by the negative sign of the second derivative of $E_{\mathbf{k}}$ with respect to k . This implies that an electron would thus be accelerated along the direction of an applied electric field, instead than opposite to it as it occurs for a free electron. Notice that no violation of Newton laws is taking place: the lattice is transferring momentum to the electron, that therefore results to behave as if it had a negative mass.

4.2.6 Introducing the scattering

4.2.6.1 Particles with mass and charge

As anticipated, expression (4.11) is not consistent with Ohm's law. This becomes apparent when the crystal periodic potential is substituted by uniform potential, as if it would result in the jellium model where ionic charges are uniformly distributed into the material, and at the same time the electron-electron interactions are being neglected. In this situation, each electron would feel only the electric field, $\hbar\mathbf{k}$ is the electron momentum and the relation (4.11) represents a Schrödinger equation for the electron in quasi-classical approximation. The solution would be given by (4.12) with no limitations on \mathbf{k} . Therefore, the current summed over all the $\mathbf{k}_i(t)$ electron trajectories

$$\mathbf{j} = -\frac{\hbar e}{m} \sum_i \mathbf{k}_i(t) \quad (4.43)$$

would be indefinitely increasing with time. This is clearly in contrast with Ohm's law. The same conclusion would be reached when starting from (4.34), taking into account that the effective mass coincides with the electron mass except for the effects due to electron-electron interactions.

The key-point here is that scattering of the carriers with other particles or excitations in the material cannot be neglected. In the following, simple phenomenological models are introduced to describe the different scattering mechanisms, postponing to the end of this Chapter a discussion of the microscopic origins of these terms.

In the phenomenological model, the scattering mechanism is described by the average time τ occurring between two scattering events. This is treated as a parameter. In a classical world, the parameter τ would naturally enter the equations of motion (4.9) or (4.11) to determine a viscous force $-m\mathbf{v}/\tau$ to be added to external forces \mathbf{F}_{ext} . Using the correspondence $m\mathbf{v} \leftrightarrow \hbar\mathbf{k}$, one would then have the phenomenological equation of motion:

$$\hbar \frac{d\mathbf{k}}{dt} = -\frac{\hbar\mathbf{k}}{\tau} + \mathbf{F}_{\text{ext}}. \quad (4.44)$$

However, the above equation is not valid in the quantum world, not even within a semiclassical approximation: in the absence of the \mathbf{F}_{ext} , the solution of (4.44) would

be $\mathbf{k}(t) = \mathbf{k}_0 e^{-t/\tau}$, leading to $\mathbf{k}(t) \rightarrow 0$ for all the electrons and violating indeed the Pauli exclusion principle. Thus,

Concept

Account of the Pauli exclusion principle requires that (4.44) be replaced by

$$\hbar \frac{d\delta\mathbf{k}}{dt} = -\hbar \frac{\delta\mathbf{k}}{\tau} + \mathbf{F}_{\text{ext}}, \quad (4.45)$$

in terms of the deviation $\delta\mathbf{k}$ of the crystal momentum from its initial value.

Expression (4.45) is strictly related to the concept of effective mass. Indeed, given the vector $\mathbf{X} = \mathbf{M} \cdot \delta\mathbf{v}$, its i -th component results to be

$$\begin{aligned} X_i &= \sum_j M_{i,j} \delta v_j = \sum_j \frac{1}{\hbar} M_{i,j} \delta \frac{\partial E_n(\mathbf{k})}{\partial k_j} \\ &= \sum_{j,l} \frac{1}{\hbar} M_{i,j} \frac{\partial^2 E_n(\mathbf{k})}{\partial k_j \partial k_l} \delta k_l = \hbar \sum_{j,l} M_{i,j} M_{j,l}^{-1} \delta k_l \\ &= \hbar \sum_l \delta_{i,l} \delta k_l = \hbar \delta k_i, \end{aligned} \quad (4.46)$$

and therefore (4.45) can be cast in the meaningful form:

$$\frac{d(\mathbf{M} \cdot \delta\mathbf{v})}{dt} = -\frac{\mathbf{M} \cdot \delta\mathbf{v}}{\tau} + \mathbf{F}_{\text{ext}}. \quad (4.47)$$

This expression replaces (4.10) when studying the motion of an electron in a periodic potential, under the action of external forces and the influence of scattering mechanisms.

The results obtained so far can be easily reconsidered in view of the modified equations of motion. In the following, the cases are illustrated when either an electric or a magnetic field is present. When only an electric field is present, the stationary solution of (4.45) results to be:

$$\delta\mathbf{k} = -\frac{e\mathbf{E}\tau}{\hbar}. \quad (4.48)$$

This represents a rigid translation of the volume $\delta\mathbf{k}$ in \mathbf{k} -space corresponding to the occupied states and bounded by the Fermi surface in the BZ, taking care that the exclusion principle be not violated. For a metal, one usually has $\delta k \ll k_f$, implying that the translation of the Fermi sphere in \mathbf{k} -space be much smaller than its radius. This point is discussed also in Sec. 4.3.5.4. This is not necessarily so in the case of semiconductors. The electric current can be calculated after relating $\hbar\delta\mathbf{k}$ to the variation of the particles crystal momentum. This can be accomplished by using, e.g., (4.46). One can get the main idea by calculating the current in the simplified

case of an isotropic effective mass, that is

$$\mathbf{j} = -en\delta\mathbf{v} = -\frac{en}{m^*}\hbar\delta\mathbf{k} = \frac{e^2n\tau}{m^*}\mathbf{E}, \quad (4.49)$$

where m^* is the effective mass at Fermi level and n the density of conduction electrons. It is manifest that the current is proportional to \mathbf{E} through the constant

$$\sigma = \frac{ne^2\tau}{m^*}, \quad (4.50)$$

which represents the electric conductivity. The inverse $\rho = \sigma^{-1}$ defines instead the electrical resistivity. If the band is completely filled, so it does also in the presence of the electric field and therefore the current $\mathbf{j} = \mathbf{0}$.

When only a magnetic field is present, the equation of motion is

$$\hbar\frac{d\delta\mathbf{k}}{dt} = -\frac{\hbar\delta\mathbf{k}}{\tau} - \frac{e}{c}\delta\mathbf{v}_{n,g} \wedge \mathbf{B}. \quad (4.51)$$

Here, the Lorentz force consistently contains the deviation $\delta\mathbf{v}$ of the velocity with respect to its initial value. Cyclotron frequencies can be consistently calculated from (4.51), with results running between two limiting cases. In the $\omega_c\tau \gg 1$ limit, an electron may wind around several closed orbits before occurring into a scattering event: therefore, all the usual topological characteristics of the constant-energy surfaces are probed and can be correspondingly investigated along the same lines followed so far. In the opposite $\omega_c\tau \ll 1$ limit, an electron scatters well before one single orbit is completed: therefore, only local information can be extracted about the surfaces with constant energy.

Quick Questions

Q11. Consider the case of a very strong electric field, so strong to trans-

late the Fermi sphere by $|\delta\mathbf{k}| > k_f$. Could the expression (4.50) for the electrical conductivity still be considered valid?

Answer - Yes, it is, provided that the effective mass be calculated at the appropriate point in \mathbf{k} -space, that is not at the Fermi level.

Q12. Softeners are molecules with positive surfactant ions, that stick to

the negatively charged fabric fibers. In this way, they leave their hydrophobic chained part to point outward: this coating gives the fabric the soft feeling. Softeners are also hygroscopic, meaning that they attract water molecules from moist air. Use the electrical properties of water to explain how the application of a softener reduces static clinging [1].

Answer - Water is a conductor, though weak: the fabric becomes slightly conducting and static electricity is therefore reduced.

4.2.6.2 Particles with spin

It is now time to introduce realistic relaxation processes for the dynamical evolution of the magnetization in (4.26) in Sec. 4.2.2.2. A simplest and reasonable assumption is that the magnetization might approach equilibrium as much faster as the actual magnetization is farther away from its equilibrium value M_0 , on a time scale T_1 characteristic of the relaxation process under consideration. Equation (4.26) would thus be replaced by

$$\frac{dM_z}{dt} = -\frac{g\mu_b}{\hbar}(\mathbf{M} \wedge \mathbf{B})_z - \frac{M_z - M_{0z}}{T_1}. \quad (4.52)$$

Typical relaxation processes involve scattering with lattice phonons. In a direct process, one phonon with energy $\hbar\omega_0$ can be absorbed or emitted, while the spin system gets to its higher or lower energy state, respectively. In a Raman process instead, one phonon can also be scattered, its energy increasing by the amount $\hbar\omega_0$ as the spin system gets to its lower state. More complicated scattering of phonons involving three spin states can also occur. In metals, absorption or emission processes of one electron can be relevant, similar to the phonon ones. In any event, the magnetization performs a precession around the magnetic field and relaxes to its equilibrium value M_{0z} on a characteristic longitudinal time T_1 .

If the magnetization components M_x and M_y transverse to $\hat{\mathbf{B}}$ are nonzero, then similar equations hold

$$\frac{dM_{x,y}}{dt} = -\frac{g\mu_b}{\hbar}(\mathbf{M} \wedge \mathbf{B})_{x,y} - \frac{M_{x,y}}{T_2} \quad (4.53)$$

for each component M_x and M_y respectively, with two differences though. First, the equilibrium magnetization $M_{0x,y} = 0$. Second, the transverse relaxation time T_2 is different from T_1 . In particular, T_2 is originated by the fact that different local fields determine different precession frequencies, so that M_x and M_y dephase with each other: T_2 is indeed a measure of such a dephasing time, typically $T_2 \ll T_1$.

Equations (4.52)-(4.53) are the so-called Bloch equations. In the event that $\mathbf{B} = B\hat{\mathbf{z}}$ and the system starts from equilibrium with $M_z = M_{0z}$, the transverse magnetizations evolve as damped harmonic oscillators: by inspection, one can indeed see that $M_{x,y} \propto e^{-t/T_2} \cos(g\mu_b B t / \hbar + \varphi)$, where after fixing for example $\varphi = 0$ for M_x , one has to take $\varphi = \pi/2$ for M_y . This result has crucial consequences and fundamental applications when a time-dependent oscillating external magnetic field is applied in the transverse direction, namely $B_x(t) = B_{0xy} \cos(\omega t)$ and $B_y(t) = -B_{0xy} \sin(\omega t)$. The situation parallels that of a damped harmonic oscillator subjected to a time-oscillating force. It is well known that after a transient, resonant oscillations set in. Their amplitude depends on the external force frequency ω according to a bell-shaped curve, whose maximum corresponds to the system natural resonance frequency ω_0 , that is $\omega = \omega_0$. In particular, the power absorption can be derived, that is

$$W(\omega) = \frac{\omega g \mu_b M_{0z} T_2 / \hbar}{1 + (\omega - \omega_0)^2 T_2^2} B_{0xy}^2. \quad (4.54)$$

Equations (4.52)-(4.53) and (4.54) are at the base of the Nuclear Magnetic Resonance method. In essence, consider to measure the frequency-dependent power absorption of a spin system subjected to a constant magnetic field along one direction and to a monochromatic oscillating magnetic field along the direction transverse to the constant magnetic field. The power absorption profile is centered at the resonance frequency ω_0 , that is a fingerprint of the nucleus that is being involved. The half-maximum width $\Delta\omega_{1/2} = 1/T_2$ of the power absorption profile is dictated by local variations of the magnetic field.

The width of the absorption profile is originated by damping or by dephasing processes. In fact, it can be driven by interaction or kinetic effects. Interaction effects can be for example originated by the interaction between two magnetic dipoles, scaling as μ/r^3 with the inverse cube of the distance. Kinetic effects can be for example originated by rapid atom motion, so that the local field values originated by e.g. dipolar interactions, are seen by a given spin to randomly fluctuate in time [12]. Each i -th time interval τ , the spin acquires a random phase change $\phi_i \approx \tau \delta E_0 / \hbar$, with $\delta E_0 = g \mu_b \delta B$ determining the linewidth $\Delta\omega_{1/2} = \delta E_0 / \hbar$ for the single event. As in a random walk, the average squared walk-variable matters, so that after n walks one has $\phi^2 = n(\tau \delta E_0 / \hbar)^2$. While large dephasings wash out, dephasings of the order of 1 radian or less contribute to the absorption line: from the expression for ϕ^2 , these can be estimated to be in the number of $n_1 \approx 1 \text{ rad}^2 / (\delta E_0 \tau / \hbar)^2$. On the other hand, $T_2 \simeq n_1 \tau$, leading to a change of the half-maximum width $\delta(\Delta\omega_{1/2}) = (\Delta\omega_{1/2})^2 \tau$. As a result, the original linewidth $\Delta\omega_{1/2}$ is reduced by a factor τ , an effect that is known as motional narrowing. By this powerful random-walk driven mechanism, the magnetic resonance linewidth of free protons can be reduced by as much as five orders of magnitude with respect to the case of protons moving around fixed lattice positions.

The position of the linewidth on the other hand, can be affected by interaction effects. One example is the hyperfine splitting, that adds to the Zeeman one, and arises from the interaction between the magnetic moments of the nucleus and of the electron. In the case of conduction electrons in metals, the shift amounts to KB with $K \approx \chi_m n_n / n$ the so-called Knight shift: the system behaves as if the paramagnetic susceptibility χ_m be modified by the ratio of the conduction electron concentration n_n close to the nucleus to the average concentration of conduction electrons n .

Therefore, while the position of the peak of the absorption profiles is a fingerprint of the atomic species under consideration, the investigation of peak shifts may infer information on the bond type and its linewidth may infer information on the interactions with neighbor spins. When performed with local scans, NMR techniques provide careful spatially-resolved images of sample composition and density profiles. The nature of the technique is also by itself not particularly invasive. These are the reasons why NMR is a so popular for medical imaging.

4.3 Metals

In this and the following 4.4-4.5 sections, the tools developed so far are exploited to investigate the transport behavior of materials with different characteristics, in regard to the phenomenon of current conduction. The equations of motion for electrons or holes in the presence of external fields and of scattering processes are here used to determine the transport behavior of metals. Along the lines explored so far, the scattering mechanisms are introduced at a phenomenological level by means of the parameter τ that works as a scattering time. Calculations of the scattering time are performed later in the present Chapter, starting from detailed knowledge of electron-ion and electron-electron interactions.

The development of the theory of conduction of metals has historically followed three fundamental steps. (i) The classical Drude theory has been first established, which helped to understand most of the basic phenomena, though showing major flaws in interpreting the experimental trends for conductivity, thermal capacitance and magnetic susceptibility as functions of temperature. (ii) The quantum Sommerfeld theory, which has repaired most of the basic flaws of Drude theory after treating the electrons as quantum particles, though still within a free-electron gas model. (iii) The band theory, which has introduced the effects of electron-ion interactions on the crystal electronic levels. In fact, only the band theory was eventually able to solve one of the most intriguing puzzles. Indeed, the setting of both the Drude model and of the Sommerfeld theory would infer that the average distance traveled by an electron between two successive scattering events, that is the mean free path, be of the order of the distance between adjacent lattice positions and also independent of temperature. The question is thus why experimental data infer instead that the mean free path is of the order of many (up to 10^{10} !) lattice distances, in fact ≈ 1 cm in very pure metals at very low temperatures. In other words, how it occurs that under suitable conditions electrons manage to move as if no other electrons and ions were present in the crystal.

Conduction in metals requires that the highest electronic band be partially empty: when subjected to the influence of external fields, electrons living in that band must have the possibility of changing their wavenumber after scattering from an occupied state with \mathbf{k} to an empty state with \mathbf{k}' in the same band. Drude and Sommerfeld had the key idea of considering electrons in metals as particles freely moving in between two scattering mechanisms. Drude was convinced that electron motion were hampered by scattering events against (fixed) ions: however, in Sec. 4.2 it has been argued that scattering against fixed ions leaves the wavevector unchanged. Scattering against lattice vibrations, impurities or any other type of topological disorder are instead those responsible of electric resistance. In essence

Concept

Ohm-like electric conduction, thermal conduction and related effects, are determined by whatever mechanism that breaks the crystal translational symmetry.

The properties of electric current in metals is now being investigated. First, consider the density of conduction electrons in metals. In the simplest approach, this is

$$n = \frac{\rho}{M} N_A Z, \quad (4.55)$$

where ρ is the material density, M the molecular weight, N_A the Avogadro number, and Z the valence of the given atomic species. Insertion of typical numbers yields electron densities of the order of $n \sim 10^{22} \text{ cm}^{-3}$ in most metals.

As to the electric conductivity, this has already been derived from (4.45) or (4.47), leading to expression (4.50) in the special case of an isotropic mass tensor \mathbf{M} .

The physical meaning of the scattering parameter τ is an additional useful information. Consider the case, typical enough, in which the scattering event lasts a very short time on the scale of the average time between two scattering events: it turns out that

Concept

Besides representing the average time between two scattering events, τ^{-1} measures the probability per unit time that an electron be scattered.

Indeed, starting from expression (4.47), one may calculate the average momentum of the electrons $\mathbf{p}(t + dt)$ at time $t + dt$, once $\mathbf{p}(t)$ is known. This is:

$$\mathbf{p}(t + dt) = \left(1 - \frac{dt}{\tau}\right) \mathbf{p}(t) + \mathbf{f}(t)dt + O(dt^2). \quad (4.56)$$

Here, $1 - dt/\tau$ represents the probability that an electron be not scattered during the time dt , and $\mathbf{f}(t)dt$ the impulse impressed by the external force $\mathbf{f}(t)$ during time dt . Neglecting terms with higher than linear order in dt , one has

$$\frac{d\mathbf{p}}{dt} \simeq -\frac{\mathbf{p}}{\tau} + \mathbf{f}(t), \quad (4.57)$$

that is the same expression as (4.45). According to this relation, τ represents also the average time between two scattering events. Indeed, the probability $P(t + dt)$ that the particle be not scattered in the time interval $(0, t + dt)$ is:

$$P(t + dt) = P(t) \left(1 - \frac{dt}{\tau}\right),$$

with $P(t)$ being the probability that it is not scattered in the interval $(0, t)$. One therefore has:

$$\frac{dP(t)}{dt} = -\frac{P(t)}{\tau},$$

that, after time integration yields

$$P(t) = P(0)e^{-t/\tau}.$$

The average time between two scattering events is by definition:

$$T = \frac{P(0) \int_0^\infty t e^{-t/\tau} dt}{P(0) \int_0^\infty e^{-t/\tau} dt} = \tau. \quad (4.58)$$

Definition

The average distance traveled by electrons between two scattering events is conveniently introduced, that is named mean-free path, $\ell = v\tau$, in terms of the average particle velocity v and of τ .

The equation of motion (4.57) can be used to calculate physical quantities that are relevant to characterize the transport behavior of metals. The general idea is that

Concept

The characteristic properties and behaviors of materials can be probed by observing the response of the system to external disturbances driving that given property. To this aim, the type and symmetry of the external disturbance is to be chosen so that it is suited to excite a fluctuation about the average value of that given property. The wavelength and frequency of the external disturbance are conveniently tuned to be on the length and time scales of the investigated fluctuation. The way the system responds, reveals information on the interactions between the particles, their statistical behavior, their effective dimensionality.

In addition,

Concept

When the external disturbance can be considered as slowly varying on the length and time scale of typical dynamical processes that are probed into the material, collision events are so frequent that each appropriately small sample volume reaches almost instantaneously conditions of local thermodynamical equilibrium. In this highly collisional, so-called hydrodynamical regime, the thermodynamical quantities such as temperature and chemical potential have a local definition, they being almost uniform across each of the small sample portions.

The concepts above are extensively used in the following, under common enough conditions in which the response of the system is linear in the external disturbance, which in turn is slowly-varying on both the time and space scale of the microscopic process under consideration.

4.3.1 Electric conductivity

Electric conductivity σ introduced in (4.50) measures the static response of a material, that under the action of an appropriate external disturbance switches on an electric current. The frequency-dependent conductivity measuring the dynamical response to rapidly varying electromagnetic fields and involved also in optical processes, is considered in Chap. 5. To drive an electric current, an electric external field can be conveniently applied. Since the wanted response is static and long-wavelength, the electric field is to be slowly-varying on the scale of the scattering time τ and of the mean-free path ℓ : that is, the frequency ω and wavenumber q of the external field are to satisfy the conditions $\omega\tau \ll 1$ and $q\ell \ll 1$. In this case, the property under investigation is represented by the current, the external disturbance by the electric field, the (static or low-frequency) response by the (static) electric conductivity.

Electric conductivity has already been calculated in (4.50) for a metal with isotropic effective masses. Notice that the current (4.49) is built up from a fluctuation $\hbar\delta\mathbf{k}$ of electron momentum. The dependence of σ on temperature T remains embodied in τ and thus not explicitly defined at this phenomenological level and until a specific microscopic representations of scattering mechanisms are used to calculate τ . Experiments show that in pure materials the resistivity $\rho = 1/\sigma$ has different power-law behaviors depending on temperature T as scaling with the Debye temperature Θ_D : linear behavior $\propto T$ as long as $T > \Theta_D$, $\propto T^5$ for $T \ll \Theta_D$, and $\simeq \text{constant}$ in the very low temperature limit $T \rightarrow 0$. This behavior is sketched in Fig. 4.8. When impurities are present and their effect is taken into account, it turns out that the resistivity can be represented as decomposed in two terms

$$\rho_{\text{tot}} = \rho_{\text{im}} + \rho(T), \quad (4.59)$$

the first ρ_{im} driven only by the concentration of impurities but independent of temperature, and the second $\rho(T)$ independent of the impurities but dependent on temperature and on the type of material. The behavior of the ρ_{im} term is displayed in Fig. 4.9 for a doped with a number of impurities [13], showing that measurements are consistent with the prediction (4.59). In essence

Concept

Resistivity in metals depends on temperature with different power-laws in different temperature ranges as in Fig. 4.8. The temperature range depends on the prevailing scattering mechanism and is thus related to the Debye Temperature. When impurities are present, resistivity shows up as if it were originated by two independent scattering mechanisms, one temperature-independent and due to impurities and one due to a temperature-dependent mechanism.

The two mechanisms being independent, the probability per unit time that an electron be scattered by either one of the two is in fact the sum of the two probabilities or else, using the physical interpretation of τ in (4.57)

Fig. 4.8 Transport properties of metals. Sketch of the temperature behavior of the electric conductivity in metals evidencing the different power-law dependence in different temperature regimes on the scale of the Debye temperature Θ_D

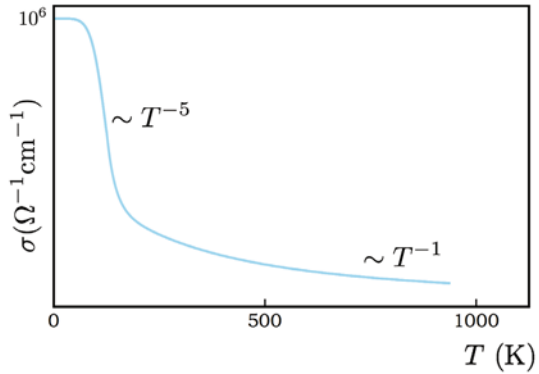
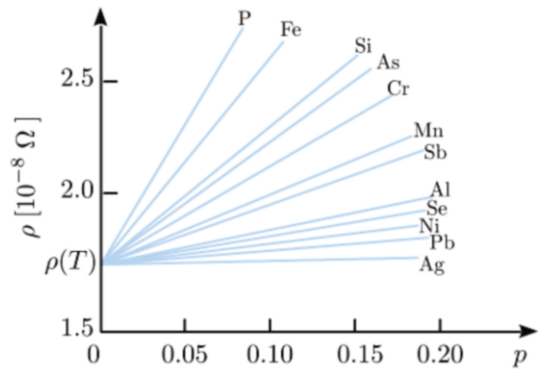


Fig. 4.9 Transport behavior of metals with impurities. Electric static conductivity response to static electric fields. The contribution of the impurity-concentration dependent term, that is the first term on the right-hand side of (4.59), to electric resistivity of Na doped with tiny concentrations p of different metallic atomic species. Different curves represent different doping species, as in the legend [13]



$$\frac{1}{\tau_{\text{tot}}} = \frac{1}{\tau_{\text{im}}} + \frac{1}{\tau}. \quad (4.60)$$

By inspection, (4.59) follows after inserting (4.60) in (4.50). Values of σ for different materials can be estimated from (4.50) once values of τ are known. Viceversa, typical values of τ can be estimated after using experimental values of σ at the given temperature. It turns out that typically $\tau \simeq 10^{-13} - 10^{-14}$ s at room temperature, also implying mean-free paths $\ell \approx 10^{-7}$ cm when the Boltzmann statistics for classical-like particles is used. When the Fermi-Dirac statistical distribution is used, mean-free paths result to be much longer as it could be predicted considering that the Pauli-exclusion principle tends to increase the electrons velocity and to keep the fermions far apart. Later on in this Chapter, the temperature dependence of resistivity in metals is tackled at a microscopic level, thus going beyond the present phenomenological description.

Quick Questions

Q13. Ultrasounds are used for medical imaging. Ultrasounds are characterized by frequencies above the audible maximum threshold of 20000 Hz and by wavelengths below the *mm* in solids, depending on the speed of sound within the medium. The working principle of ultrasounds is to send on tissues sound waves produced by means of a transducer, and measure by the same transducer the part of the wave that is reflected (echoed) back because of different impedance that different tissues offer to sound propagation. Usually, ultrasounds of deeper organs like liver or kidneys are less resolved than those of muscles or breast tissues. Argue why [1].

Answer - Liver and kidneys are deeper inside the body than muscles or breast tissues: longer wavelength are therefore better used for the exciting ultrasound, thereby loosing in spatial resolution.

Q14. Imagine to switch on an energy current \mathbf{j}_e . Which type of external disturbance could be suited to the purpose? In analogy with equations (4.49) and (4.50) and concepts discussed so far, write a phenomenological equation relating the energy current and its generating field.

Answer - An energy flow can be generally switched on by temperature differences. To be more precise, a temperature gradient ∇T . Labeling χ to be the coefficient representing the response of the system, for small temperature gradients one could assume $\mathbf{j}_e = \chi \nabla T$. Such a situation is discussed in detail in Sec. 4.3.2.

Q15. Two among possible scattering mechanisms are represented in (4.60). Indicate at least one more scattering mechanism.

Answer - As already discussed, scattering mechanisms leading to ohmic conductivity are all those breaking the crystal periodicity. Along this view, scattering against surfaces confining the sample, or else, scattering with other electrons are additional example possibilities.

4.3.2 Thermal conductivity and thermoelectric effects

Thermal conductivity χ_c and thermoelectric coefficient χ_t measure the response of a material that switches on an energy current and an electromotive force, respectively, under the action of a suited external disturbance. To excite an energy current, a temperature gradient is to be applied. One could also imagine to apply a pressure gradient as well: however, as it is discussed in Chap. 6, this would violate momentum conservation and is therefore impossible. Once again, the response of the system is considered at low frequencies and long wavelengths. Thermoelectric effects result

in the occurrence of an electromotive force under conditions of vanishing electron currents, once a temperature gradient has been applied.

Consider indeed a metal bar, whose ends be in contact with two heat sources at different temperatures. Experience shows that, after a transient time, stationary conditions set in: temperature gradually changes from one end to the other, each transverse section of the bar being crossed by both an energy and an electric current. In addition, if the electric current is required to vanish, as e.g. if the circuit including the bar is left open, then an energy current yet sets in, though an electric field appears inside the material. In mathematical language, as a consequence of an applied temperature gradient ∇T , an energy current \mathbf{j}_e manifests according to the so-called Fourier law

$$\mathbf{j}_e = -\chi_c \nabla T, \quad (4.61)$$

as well as an electric field \mathbf{E}_t

$$\mathbf{E}_t = -\chi_t \nabla T. \quad (4.62)$$

In the present regime of weak temperature gradients, both the energy current \mathbf{j}_e and thermoelectric field \mathbf{E}_t are linearly dependent on the driving external field ∇T , according to the respective response coefficients characteristics of the given material: the thermal conductivity χ_c and the thermoelectric coefficient χ_t . While χ_c significantly depends on the material, this is not the case for χ_t . As to their temperature dependence, χ_c is almost independent of T at large temperatures $T \gg \Theta_D$, is $\propto T^{-2}$ for $T \ll \Theta_D$ and $\propto T$ for $T \rightarrow 0$, as sketched in Fig. 4.10. In general, χ_t is linearly dependent on T .

In order to help fixing the key concepts, simple though physically meaningful expressions for the two responses are now being derived within a simple one-dimensional model and in terms of phenomenological parameters. Consider a metal bar, whose ends be in contact with two heat sources at different temperatures. Select a transverse section of the bar between the two ends, across which electrons move in both directions with different energies and velocities. The following two model approximations help simplifying the treatment, while leaving the possibility of catching the basic physics. First, local thermodynamical equilibrium is assumed: that is to say that the material can be thought as a collection of a large number of volumes. The size of these elementary volumes is to be sufficiently small that their temperature be considered spatially uniform, and at the same time large enough to contain a sufficiently large number of particles to determine thermodynamic equilibrium by means of statistical calculations. Second, thermodynamical equilibrium conditions within each volume are reached in almost instantaneous manner: that is, each time an electron undergoes scattering events within the above volume, its energy distribution is consistent with the equilibrium thermodynamical conditions appropriate to that portion of material.

For a quantitative analysis, an electric field \mathbf{E} is also conveniently introduced in addition to the temperature gradient and parallel to it. The reasons for introducing \mathbf{E} will become clear later on. Each electron is thus expected to acquire a drift velocity $\mathbf{v}_d = -e\tau\mathbf{E}/m$ due to the electric field, which sums up to the velocity due to thermal effects. Therefore, given two electrons crossing the material section along

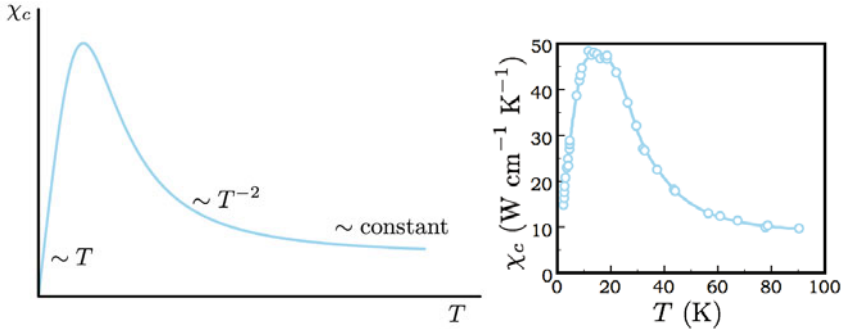


Fig. 4.10 Transport behavior of metals. Thermal static conductivity response to static temperature gradients. Left panel: sketch of the temperature behavior in metals evidencing the different power-law dependence in different temperature regimes on the scale of the Debye temperature Θ_D . Right panel: experimental data for copper [14]

$\hat{\mathbf{x}}$ on opposite sides, they have velocities $v_d + v_x(x - \ell_x)$ and $v_d - v_x(x + \ell_x)$. Here, $v_x(x - \ell_x)$ and $v_x(x + \ell_x)$ are the x -components of the average velocities that the electrons possess in thermodynamical equilibrium at the positions $x - \ell_x$ and $x + \ell_x$, where they experienced the last scattering event characterized by a mean-free path ℓ_x . Be n the (uniform) volume density of electrons at $x - \ell_x$ and $x + \ell_x$ and $\varepsilon(x)$ the average energy of each particle. Under stationary conditions, the energy current j_e and the electric current j result to be:

$$j_e = \frac{n}{2}[v_d + v_x(x - \ell_x)]\varepsilon(x - \ell_x) + \frac{n}{2}[v_d - v_x(x + \ell_x)]\varepsilon(x + \ell_x)$$

$$j = -\frac{n}{2}e[v_d + v_x(x - \ell_x)] - \frac{n}{2}e[v_d - v_x(x + \ell_x)]. \quad (4.63)$$

Expand now energy $\varepsilon(x)$ and velocity v_x in series to first order in ℓ_x , under the assumed conditions of vanishing average electric current. One thus obtains:

$$j_e = -v_x n \ell_x \frac{d\varepsilon}{dx} = -n \ell_x v_x \frac{d\varepsilon}{dT} \frac{dT}{dx} = -n \tau v_x^2 \frac{d\varepsilon}{dT} \frac{dT}{dx}$$

$$= -n \frac{v^2}{3} \tau \frac{d\varepsilon}{dT} \frac{dT}{dx} = -\frac{v^2}{3} \tau n c_V \frac{dT}{dx}$$

$$env_d = n \ell_x e \frac{dv_x}{dT} \frac{dT}{dx} = \frac{n \tau e}{2} \frac{dv_x^2}{dT} \frac{dT}{dx}$$

$$= \frac{n \tau e}{6} \frac{dv^2}{dT} \frac{dT}{dx} = \frac{\tau e}{3m} n c_V \frac{dT}{dx}, \quad (4.64)$$

where a three-dimensional average $v_x^2 = v^2/3$ has now been performed. Notice that in (4.64), j_e and env_d are entirely expressed in terms of two characteristic system properties besides charge and mass, that are the specific heat c_V per particle and of the mean-free path $\ell_x = v_x \tau$, as discussed also in Problem 4.2. Substituting v_d on

the left-hand side of the last equation with $-\sigma E/(en)$, one finds that

$$E = -\frac{c_V}{3e} \frac{dT}{dx}. \quad (4.65)$$

Thus, the artificially introduced electric field E is precisely what is needed to ensure the conditions of open circuit and vanishing electric current: E can therefore be identified with the thermoelectric field E_t that manifests itself inside the material if the average current is to be zero.

In conclusion, one has the following expressions for the two response coefficients:

$$\begin{aligned} \chi_c &= \frac{v^2}{3} \tau n c_V \\ \chi_t &= \frac{c_V}{3e}. \end{aligned} \quad (4.66)$$

Notice that

Concept

The temperature dependence of χ_c and χ_t follows the T -dependence of τ and c_V .

In particular, the dependence of χ_t on temperature is the same as the dependence of c_V on it. Once again, the T -dependence is determined by the statistics that is used to calculate averages, whether it is classical Boltzmann or quantum Fermi-Dirac or Bose-Einstein statistics. Experimental data are consistent with the use of quantum statistics, as it is discussed later on in this Chapter.

The first between the two equations (4.66) can be extended to any type of energy carriers different from electrons, they possibly being positively-charged or neutral particles, or else crystal excitations. In metals for example, thermal conduction is contributed by both electron dynamics and by ion vibrations, that is phonons, that transfer energy from hotter to colder sample portions.

Concept

In metals, electron contributions to thermal conductivity are more significant than phonon ones, and typical thermal conductivities are one or two orders of magnitude higher than those in insulators. In insulators, phonon-driven thermal conductivity is clearly the only possibility. In semiconductors, either one of the two contributions can be predominant, strongly depending on the number of impurities, which the material is doped with.

Quick Questions

Q16. Is the thermoelectric effect possible in insulators?

Answer - No, it is not. The second equation in (4.66) intrinsically requires the existence of an electric current, thus the effect may occur in metals and semi-conductors.

Q17. Consider a slab of metallic material having an approximately two-dimensional geometry. Is the dependence of χ_c and χ_t on τ and c_V still valid as in (4.66)?

Answer - Yes, it is, though a few factors change as a result of 2D instead of 3D-performed averages. The two expressions are indeed turned into:

$$\begin{aligned}\chi_c &= \frac{v^2}{2} \tau n c_V \\ \chi_t &= \frac{c_V}{2e},\end{aligned}\tag{4.67}$$

where c_V is the specific heat per particle as calculated in 2D. In particular, the factor 2 comes about after performing a planar angular average of v_x^2 , in place of the factor 3 in (4.66) that results from the average over a solid angle .

Q18. Our body is able to keep thermal energy. Guess why.

Answer - Skin is composed of substances with low thermal conductivity: fat for example, has a thermal conductivity that is 20% about that of water and only 0.03% that of copper. Skin is sufficiently thick to disfavor thermal energy flow towards the environment (not too thick for practical reasons) and the shape of our body is such as to minimize its surface [1] .

Q19. You feel cold when touching a metallic spoon. Why?

Answer - A metallic spoon is a good thermal conductor. While touching it, thermal energy flows from your hand to the spoon, providing the feeling of coldness [1] .

Q20. You must caulk the roof of your house. The constructor proposes

to use a special foam which the roof should be filled in. Is he/she right? [1]

Answer - Yes. Foams contain air bubbles and air is a poor thermal conductor .

4.3.3 *Transport properties of metals within the Drude model*

This section discusses the temperature dependence of transport properties of metals, as they can be determined within the Drude model. Comparison of the found behavior with known experimental data claims the inadequacy of Drude model and suggests the way to improve it. In the Drude model to describe the transport behavior of metals, the particles undergo scattering events, and in between each two of them they evolve as if they were free particles. The scattering processes can be assumed to leave the total energy almost unchanged. On the other hand, the energy of the free motion coincide by definition with the kinetic energy. Within the Drude model, the particles are assumed to behave in a classical manner, that is their energy is assumed to be distributed according to the Boltzmann statistics. Equipartition theorem assigns an average energy $k_b T/2$ to each degree of freedom that appears in a quadratic form in the total energy. Thus, the statistical averages leading to expressions for the different transport variables and their dependence on temperature can be easily performed, some of them requiring also the temperature behavior of τ . Comparison of the results with experimental data evidences the need to consider the quantum nature of the carriers participating to the transport processes. As tackled in the next Sec. 4.3.4, this requires the introduction of quantum statistics, in particular the Fermi-Dirac distribution for electrons, as more refined Sommerfeld theory of metals does. With this order of presentation, mathematical difficulties ramp up slowly: the derivation of equations is simpler in the Drude model and can be easily complemented at a second time with the modifications due to the Sommerfeld theory.

4.3.3.1 Conductivity

The original Drude's idea to study electric conduction and related phenomena in crystals, was to apply the classical kinetic gas theory to an ensemble of charged particle, namely electrons. The simple assumptions which the Drude model relies on can be summarized as follows.

(i) The electrons behave as if they were a gas of classical particles, but with no collisions: no correlations are included, originated neither from quantum statistics nor from Coulomb interactions. This is the content of the independent electron approximation.

(ii) Electrons can collide with the ion cores, which are though assumed to occupy fixed positions. Each collision determines an instantaneous change of electron velocity. After each collision the memory of the previous event is completely lost, that is the velocity of the electron emerging after a collision is fully randomized. Between each two collisional events, the electron motion is completely free, that is no electron-ion interactions take place. This is the content of the free-electron approximation.

(iii) The electrons are in thermal equilibrium with the ion lattice. According to the classical kinetic gas theory, the average kinetic energy at temperature T is given by

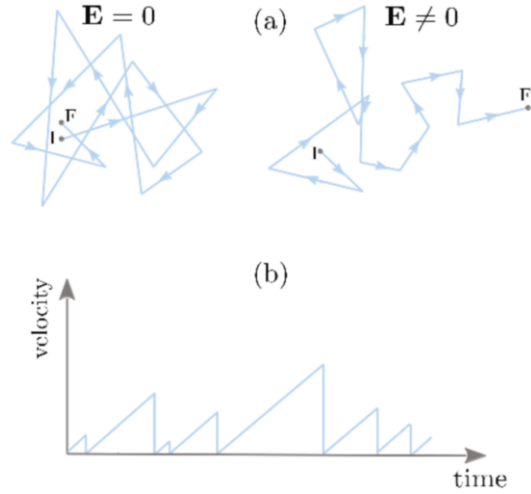


Fig. 4.11 Transport behavior in metals. Drude model for electric conduction in crystals. (a) Illustration of the electron motion between initial I and final F points. The concept of drift velocity is sketched for both the zero (left) and non-zero (right) electric field. (b) The change of drift velocity with time in the presence of an electric field

the equipartition theorem, that is

$$\frac{1}{2}m_e v_t^2 = \frac{3}{2}k_b T.$$

At room temperature, this corresponds to an average thermal velocity $v_t \approx 10^5 \text{ ms}^{-1}$. In the absence of any external perturbation, it is evident that the electron motion is completely random. This means that the averaged vectorial thermal velocity over the whole electron gas is zero, that is:

$$\bar{\mathbf{v}}_t = \frac{1}{N} \sum_{i=1}^N \mathbf{v}_{t,i} = \mathbf{0},$$

with $\mathbf{v}_{t,i}$ the thermal velocity of one single electron. Therefore, no net flow of electrons takes place, corresponding to a zero electric current.

(iv) As discussed in Sec. 4.2.6, the probability $1/\tau$ per unit time for the electrons to undergo a scattering event is introduced. The relaxation time τ corresponds to the average time between two collision events, and the mean free path can be defined as $\ell = v_t \tau$. Being the ions considered fixed to their equilibrium positions, typical interatomic distances $\ell \approx 1 \text{ nm}$ lead to typical values $\tau \approx 1 \cdot 10^{-14} \text{ s}$ for the relaxation time. According to the Drude model, in the absence of electric fields an electron does not drift its average position, despite its random motion depicted in the left panel of Fig. 4.11(a). After application of an external electric field, the electron starts drifting along the field direction: a net charge flow along the field is superimposed to the random thermal motion and an electric current sets in. A picture of this latter motion is illustrated in Fig. 4.11(b). According to the results of Sec. 4.2.6, in the absence of any scattering event the drift velocity should linearly increase with time, in contradiction with experimental evidences. However, under the assumption that the

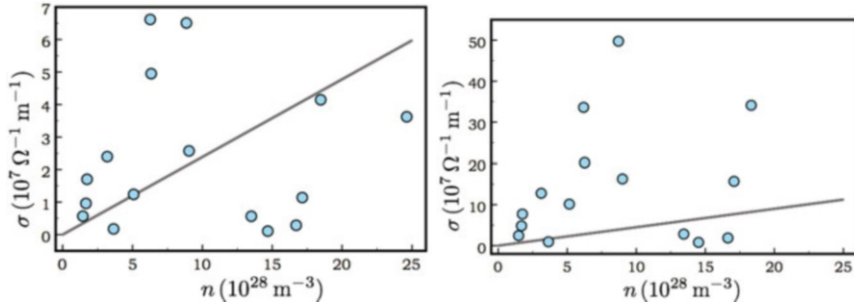


Fig. 4.12 Transport behavior of metals. Electric conductivity. The experimental measurements for a number of metals (dots) are compared with the theoretical prediction of the Drude model (straight lines) at two different temperatures. Left panel: $T = 273$ K. Right panel: $T = 77$ K. The conductivity values are reported as a function of the measured electron density

electron undergoes scattering against the ion cores, after each collision the electron motion is randomized and the drift velocity resets. This is illustrated in Fig. 4.11(b).

Notice that the assumptions of Drude model lead to expect that the mean free path ℓ be independent of temperature and with size of the order of typical inter-atomic distances. On the other hand, the electron travels between each two scattering events with thermal velocity $v_t \propto 1/\sqrt{T}$. As a result, $\tau = \ell/v_t \propto 1/\sqrt{T}$ and finally $\sigma \propto 1/\sqrt{T}$ from (4.50). Fig. 4.12 evidences the comparison between experimental and Drude theoretical conductivities for selected metals. While some degree of compatibility shows up at higher temperatures with the experimental data scattered around the theoretical prediction, at lower temperatures not even the general trend of the experimental data is correctly reproduced by the Drude prediction.

One more shortcoming of the Drude model is given by the impurity contribution to the resistivity, as shown in (4.59). Indeed, there is no apparent reason why the presence of impurities should produce a drastic decrease of the conductivity: an impurity can indeed be considered a scattering center just as the ions do are in their fixed equilibrium positions.

4.3.3.2 Specific heat

The discussion starts by calculating the electronic contribution to specific heat entering some of the transport properties investigated in Sec. 4.3.1. This derivation parallels the one performed in Chap. 3 to determine the phonon contribution.

Consider all the assumptions of the Drude model discussed above. Using equipartition theorem and Boltzmann statistics, the energy density of a gas of free electrons and the corresponding specific heat per particle are

$$U = \frac{3}{2}nk_bT$$

$$c_V = \frac{1}{n} \frac{\partial U}{\partial T} = \frac{3}{2} k_b, \quad (4.68)$$

the factor of 3 in U accounting for the three degrees of freedom of one particle with just kinetic energy moving in 3D space. In facts, experimental data show that the specific heat of metals is a linearly increasing function of temperature, the proportionality coefficient being a few hundredths of $3k_b/2$ at room temperature. This is clearly in contrast with (4.68). Moreover, even at room temperature the experimental findings follow the Dulong-Petit $3R$ value of the specific heat for a metal, as if only the lattice were contributing and the electronic term were either absent or negligible. Actually, the electronic contribution to the specific heat is usually not larger than 1% of the value predicted by (4.68). Therefore, the system behaves as if the number of electrons contributing to the electric conductivity were only about 1/100 of the number N contributing to the thermal capacity. This tile of the puzzle is put in place within the quantum description of the electron gas.

Quick Questions

Q21. Consider a metal in 2D geometry. How is the expression (4.68) changed ?

Answer - One has

$$\begin{aligned} U &= nk_b T \\ c_V &= \frac{1}{n} \frac{\partial U}{\partial T} = k_b, \end{aligned} \quad (4.69)$$

since two degrees of freedom contribute to U instead of 3. .

4.3.3.3 Wiedemann and Franz law

The dependence of σ and χ_c conductivities on temperature requires the knowledge of the function $\tau(T)$. On the other hand, both conductivities are proportional to τ and therefore their ratio

$$\frac{\chi_c}{\sigma} = \frac{v^2}{3} \tau n c_V \frac{m}{ne^2 \tau} \quad (4.70)$$

does not. Inserting the expressions for c_V and v^2 within the Boltzmann statistics, the temperature dependence only comes from v^2 and one finds

$$\frac{\chi_c}{\sigma T} = \frac{3k_b^2}{2e^2} = 1.11 \cdot 10^{-8} \frac{\text{W}\Omega}{\text{K}^2}. \quad (4.71)$$

This result is known as Wiedemann and Franz law:

Concept

The ratio of electric-to-thermal conductivity does not depend on the scattering mechanism, the scattering time τ not appearing in it.

This result is mostly well verified by experiments at temperatures $T > \Theta_D$, except for a factor ~ 2 on the right-hand side of (4.71). For $T < \Theta_D$ instead, $\chi_c/(\sigma T) \propto T^2$. The Wiedemann and Franz law is the best success of the Drude model, though the temperature range of validity remains unclear.

Quick Questions

Q22. Consider a metal in a 2D geometry. Would the Wiedemann and Franz law be valid?

Answer - Yes, it is, but the value of the constant would be k_b^2/e^2 .

4.3.3.4 Magnetoresistance and classical Hall effect

In this paragraph the effects on electric transport behavior are analyzed, when a magnetic field is present as well. Reverting back to (4.30), it has been shown that an electric current \mathbf{j}_\perp perpendicular to both fields takes place in the material, under the action of time-independent electric and magnetic fields orthogonal to each other, say $\mathbf{E} \equiv (E_x, E_y, 0)$ and $\mathbf{B} \equiv B_z \hat{\mathbf{z}}$.

This result can also be viewed in terms of the response of the system to static electric and magnetic fields that are perpendicular to each other, leading to a current j_x along the direction perpendicular to both fields. In the $\omega_c \tau \gg 1$ limit of large magnetic fields B_z in $\omega_c = eB_z/(mc)$ or long scattering times τ , the current density is $j_x = nec(E_y/B_z)$. The peculiar combination of $x-y-z$ components of the different vectors in determining the system response suggests that a more complicated than vectorial representation is here involved, which is better understood later on.

The quantity $R_H^{-1} \equiv j_x B_z / E_y$ represents the response of the system to such an external disturbance. From (4.30), one has

$$R_H = -\frac{1}{nec} \quad (4.72)$$

in terms of intrinsic properties of the charge carriers composing the current. This effect is named after Hall. Deviations from this result occur for both the sign of the Hall coefficient R_H and for its dependence on the magnetic field intensity.

Assume to have a material sample that is confined to a finite size in z -direction, and apply electric field $(E_x, E_y, 0)$ and magnetic $(0, 0, B_z \equiv B)$ fields. The semiclassical equation of motion for an electron in the material is

$$\frac{d\delta\mathbf{p}}{dt} = -e \left(\mathbf{E} + \frac{\delta\mathbf{v}}{c} \wedge \mathbf{B} \right) - \frac{\delta\mathbf{p}}{\tau}, \quad (4.73)$$

or else, decomposed along the x, y and z directions:

$$\begin{aligned} \frac{d\delta p_x}{dt} &= -e \left(E_x + \frac{\delta v_y}{c} B \right) - \frac{\delta p_x}{\tau} = -e \left(E_x + \frac{\delta p_y}{mc} B \right) - \frac{\delta p_x}{\tau} \\ \frac{d\delta p_y}{dt} &= -e \left(E_y - \frac{\delta v_x}{c} B \right) - \frac{\delta p_y}{\tau} = -e \left(E_y - \frac{\delta p_x}{mc} B \right) - \frac{\delta p_y}{\tau} \\ \frac{d\delta p_z}{dt} &= -\frac{\delta p_z}{\tau}. \end{aligned} \quad (4.74)$$

The third relation admits the solution $\delta p_z(t) = \delta p_{0,z} e^{-t/\tau}$, that is vanishing at long times $t \gg \tau$: the particles indeed move onto planes that are perpendicular to \mathbf{B} . The remaining two equations are more easily solved after introducing the complex function of real variables $u(t) = \delta p_x(t) + i\delta p_y(t)$ along with $\tilde{E} = E_x + iE_y$. Insertion of this function into the two equations yields:

$$\frac{du}{dt} = -\frac{u}{\tau} - e\tilde{E} + \frac{eB}{mc}iu = -u \left(\frac{1}{\tau} - i\omega_c \right) - e\tilde{E}, \quad (4.75)$$

with $\omega_c = eB/(mc)$. Equation (4.75) has been obtained after suited summation of the former two equations (4.74), and admits the solution:

$$u(t) = u(0)e^{-\frac{t}{\tau}}e^{-i\omega_c t} - \frac{e\tilde{E}\tau}{1 - i\omega_c\tau}. \quad (4.76)$$

After a transient time thus, the electric currents along the x and y axes are:

$$\begin{aligned} j_x &= \frac{\sigma}{1 + (\omega_c\tau)^2} (E_x - \omega_c\tau E_y) \\ j_y &= \frac{\sigma}{1 + (\omega_c\tau)^2} (\omega_c\tau E_x + E_y). \end{aligned} \quad (4.77)$$

Equations (4.77) remark that current and electric field are connected by a tensorial relation. That is,

Concept

The current response $\mathbf{j} = \boldsymbol{\sigma}_H \cdot \mathbf{E}$ to a combined electric and magnetic orthogonal fields is represented by a tensor $\boldsymbol{\sigma}_H$, so that electric currents in a given direction set in with intensity determined by the electric field in a different direction: the magnetic field is responsible for the off-diagonal terms. $\boldsymbol{\sigma}_H$ generalizes the concept of electric conductivity σ and is named magnetoconductivity.

The magnetoconductivity components are $\sigma_{xx} = \sigma_{yy} = \sigma[1 + (\omega_c\tau)^2]^{-1}$ and $\sigma_{xy} = -\sigma_{yx} = -\sigma\omega_c\tau[1 + (\omega_c\tau)^2]^{-1}$. Notice that in the absence of a magnetic field,

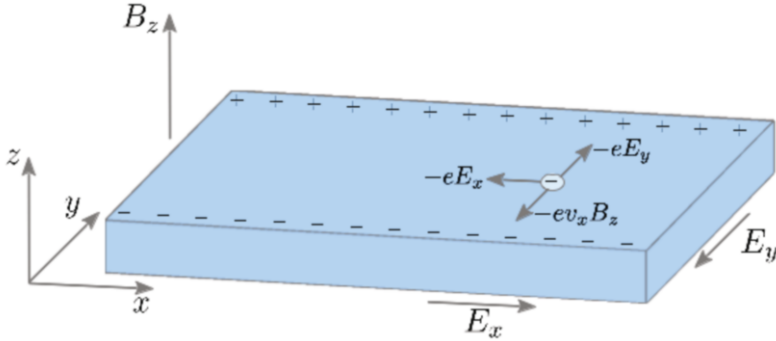


Fig. 4.13 Transport behavior in systems with reduced dimensionality and in the presence of magnetic fields. Schematic illustration of the Hall effect

$\omega_c = 0$, the off-diagonal terms of σ_H vanish and the remaining diagonal terms of (4.77) recover the electric conductivity response relation. In essence, the magnetic field originates a sort of system anisotropy with the appearance of an electric current, say j_x (j_y), in direction orthogonal to the electric field E_y (E_x).

The significance of such off-diagonal terms is driven by $\omega_c \tau$. In the strongly collisional regime with $\omega_c \tau \ll 1$, the average time between two scattering events is much shorter than the cyclotron period $1/\omega_c$: so many scattering events occur in $1/\omega_c$, that the system somehow loses any memory of the presence of \mathbf{B} and the usual conductivity response result $\mathbf{j} = \sigma \mathbf{E}$ is recovered from (4.77). If $\omega_c \tau \gg 1$ instead, the electrons winds many times around before incurring into a scattering, and the off-diagonal terms in (4.77) become larger than the diagonal ones.

Consider now the case in which the sample is confined as well along the y direction, and the electric field is initially directed only along x direction. A y -directed current thus sets in, leading to accumulation of charges of opposite sign on the two opposite surfaces limiting the sample along y . This separation of charges originates an electric field along y , whose intensity increases during the charge accumulation process. A schematic picture is given in Fig. 4.13.

This situation lasts until $j_y = 0$, that is

$$E_y = -\tau \omega_c E_x. \quad (4.78)$$

Under these conditions, the current in x direction amounts to

$$j_x = ne^2 \tau E_x / m, \quad (4.79)$$

restoring the same result obtained in the absence of a magnetic field. Expressions (4.78)-(4.79) lead to:

$$E_y = -\frac{B}{nec} j_x. \quad (4.80)$$

Therefore, a measure of the Hall coefficient

$$R_H = \frac{E_y}{j_x B} = -\frac{eB\tau}{mc} \frac{E_x}{ne^2\tau E_x/m} = -\frac{1}{nec}, \quad (4.81)$$

under these conditions would provide information on the electronic density, (4.81) being independent of τ and any other material-dependent property. Experiments show strong deviations from this formula. The latter is recovered only in the limit of strong values of the magnetic field.

Quick Questions

Q23. Consider the situation in which electric and magnetic field have components $\mathbf{E} \equiv (E_x, 0, 0)$ and $\mathbf{B} \equiv (0, 0, B)$. Discuss the conditions for j_x be larger than j_y .

Answer - From (4.77) one has

$$\frac{j_x}{j_y} = \frac{1}{\omega_c \tau}.$$

Thus $j_x/j_y > 1$ for weak B fields with $\omega_c \tau < 1$, whereas $j_x/j_y < 1$ for strong B fields with $\omega_c \tau > 1$.

For each selected compound, the expected value of the Hall coefficient can be calculated from Eqs. (4.81) and (4.55).

Table 4.1 lists the ratio of the measured R_H at very large values of magnetic field ($> 10^4$ G), to the calculated R_H^c Hall coefficient for selected metals. The expression for $R_H^c = -(nec)^{-1}$ is calculated after inserting the electron density n (4.55) based on the nominal chemical valence Z . If this expression were applicable, the expected ratio would be 1 for all compounds. It turns out that modest deviations are observed for alkali metals and larger, though still acceptable, deviations for noble metals. In all the other cases, the deviations are significant at both a quantitative level, with very large deviations from 1, and a qualitative level, with an inverted sign of the coefficient. Bi represents a striking example of quantitative discrepancies. Be, Mg, In and Al represent striking examples of the above qualitative discrepancy, as if the conduction were driven by carriers with charge of opposite sign than that of electrons. A look at the data in third column suggests indeed that the given compound behaves as if it has an effective chemical valence of $n_v = R_H/R_H^c$. Along these lines, Be would behave as if it had 0.4 carriers per primitive cell and with a positive charge, Mg approximately 0.8 positive carriers per primitive cell, In and Al approximately 1. Bi would behave as if it had an infinitesimally small valence. All these discrepancies cannot be explained within the Drude model.

Table 4.1 Hall effect in selected metals (from [2]). Ratio R_H/R_H^c of the measured R_H at very large values of magnetic field ($> 10^4$ G), to the calculated R_H^c Hall coefficient $R_H^c = -(nec)^{-1}$. The predicted $R_H^c = -(nec)^{-1}$ is calculated using the free electron density n based on the nominal chemical valence Z recalled in the second column. The data in third column suggest that the compound behaves as if it had an effective valence number $n_v = R_H/R_H^c$.

| Element | Z | R_H/R_H^c |
|---------|-----|-------------------|
| Li | 1 | 0.8 |
| Na | 1 | 1.2 |
| K | 1 | 1.2 |
| Rb | 1 | 1.0 |
| Cs | 1 | 0.9 |
| Cu | 1 | 1.5 |
| Ag | 1 | 1.3 |
| Au | 1 | 1.5 |
| Be | 2 | -0.2 |
| Mg | 2 | -0.4 |
| In | 3 | -0.3 |
| Al | 3 | -0.3 |
| Bi | 5 | $\approx 1/40000$ |

4.3.3.5 Paramagnetic susceptibility and Pauli paramagnetism

In Chap. 2, the manifestation of magnetic effects in materials after application of an external magnetic field has been introduced. Materials have been classified as diamagnetic, paramagnetic, and ferro-, ferri- and antiferromagnetic. In essence, electrons in atoms or molecules composing the material produce small magnetic fields as they spin and move around the atom. If ℓ is the spin quantum number, they possess an intrinsic magnetic dipole moment $\mu = 2\mu_b\sqrt{\ell(\ell + 1)}$ with μ_b the Bohr magneton: its component $\mu_z = 2\mu_b s_z$ along the field is proportional to the eigenvalue of the component s_z of the spin in z direction, apart from a factor \hbar . When electrons possess both orbital angular momentum and spin, the quantum numbers for total angular momentum j and m_j are instead involved and from quantum mechanics calculations one has $\mu = \mu_b g \sqrt{j(j + 1)}$ and $\mu_z = \mu_b g m_j$, with g the Landé factor. The angular momentum and spin of nuclei should also be added, though the heavier mass of nuclei appearing in the Bohr magneton suggests that their magnitude can be neglected in the present context.

Now, in diamagnetic materials, these atomic magnetic moment cancel out when the field is off, but a magnetic moment is induced after the application of the magnetic field, which the induced magnetic moment opposes against. In paramagnetic materials instead, a finite intrinsic magnetic moment is present from the beginning: this is then oriented by the applied magnetic field so to produce an additional magnetization, the latter being dependent on the orientation statistical distribution and thus on temperature. Ferromagnetic and antiferromagnetic materials are those in which below a given critical temperature a new ground state is energetically favorite, in

which a macroscopic alignment of magnetic moments permanently manifests even in the absence of the magnetic field: with parallel alignment in ferromagnets so to have a finite macroscopic permanent magnetic moment, antiparallel with different sizes and nonzero net magnetic moment in ferrimagnets, and antiparallel with equal sizes and zero net magnetic moment in antiferromagnet.

In this context, magnetic susceptibilities measure the response of the material to the presence of an applied magnetic field: while diamagnetic susceptibilities are negative, paramagnetic ones are positive. In both cases magnetic susceptibilities are usually very small $\ll 1$, whereas they are very large in (anti)ferromagnetic materials.

Considered the classification above, all insulators are diamagnetic while metals and semiconductors show up both paramagnetic and diamagnetic behavior. In Chap. 2, it has been shown that the paramagnetic susceptibility χ of N atoms arranged in a crystal follows the $1/T$ Curie law $\chi \simeq Ng^2\mu_b^2s(s+1)/(3k_bT)$: here the spin quantum number s appears in place of the total j entering χ of one single atom, since atoms are arranged in a crystal and j is not any longer a good quantum number. The paramagnetism of conduction electrons, called Pauli paramagnetism, remains now to be considered.

To this aim, it is first useful to focus the main point here. In Chap. 2, the explicit calculation of the paramagnetic susceptibility has led to the result (2.209), that is

$$\chi_m = \frac{n\mu_s^2}{3k_bT}. \quad (4.82)$$

Now, experimental data on the paramagnetism of conduction electrons show instead that χ_m in metals is essentially independent of T : this is a major flaw of the Drude model. Overcoming the discrepancies between the Drude model and the experimental facts requires a qualitative change of perspective: this task is pursued below, exploiting all the tools developed so far.

4.3.4 Comparison with experiments requires Fermi-Dirac statistics: the Sommerfeld theory

Drude model shows up to be inadequate to describe the experimental behavior of the system responses to external fields, such as electric and thermal conductivity, specific heats and Pauli paramagnetic susceptibility, both qualitatively in terms of temperature dependence and quantitatively in terms of the size of given physical quantities. Besides, the microscopic knowledge of the scattering time τ yet remains to be uncovered.

The theory of metals due to Sommerfeld repairs many of these basic flaws. In essence, Sommerfeld theory still consider the charge carriers contributing to the transport behavior as free particles, but takes into account their quantum statistical nature. In particular, classical Boltzmann statics is replaced with quantum Fermi-

Dirac statistics in the calculation of thermodynamical averages, and Newton equations are replaced by Schrödinger equations when determining the particle motion.

The relevant quantities and parameters describing the gas of independent quantum fermionic particles are usefully recalled here from Chap. 2. The density of states $g(\varepsilon)$ of free electrons per unit volume was found to be $g(\varepsilon) = \sqrt{2}m^{3/2}(\pi^2\hbar^3)^{-1}\sqrt{\varepsilon}$. At $T = 0$ K, the highest energy of the occupied states is the Fermi energy $\varepsilon_f = \hbar^2 k_f^2/(2m)$, in terms of the Fermi wavevector $k_f = (3\pi^2 n)^{1/3}$ and of the density n . Typical values of k_f in metals are $k_f \sim 10^8 \text{ cm}^{-1}$, that is the same order of magnitude as the BZ size. Typical values of ε_f in metals are $\varepsilon_f \equiv k_b T_f \sim 1 \text{ eV}$, so that typical Fermi temperatures are of the order of $T_f \simeq 10^4 \text{ K}$. As a result, typical values of the velocities of electrons at the Fermi level are $v_f = \hbar k_f/m$, of the order of $\simeq 10^8 \text{ cm/s}$. Notice that k_f values are comparable with the BZ size: this supports a posteriori the validity and use of the semiclassical approximation to describe the electron motion under the action of external fields. Thus, (4.45) or (4.47) are still valid: the description of the electrons dynamical behavior is formally similar to that developed within the Drude model, except for the substantial different statistics that is being used. In particular, the calculation of thermodynamic averages involves the Fermi-Dirac occupation function

$$f(\varepsilon) = \frac{1}{e^{\beta(\varepsilon-\mu)} + 1},$$

with $\beta \equiv (k_b T)^{-1}$. Indeed, the classical Boltzmann distribution applies only to distinguishable particles and predicts that a system with N free electrons at zero temperature has zero energy: as a result, all particles would have zero kinetic energy and, being non-interacting, zero total energy. But this is not consistent with quantum mechanics: according to the Pauli exclusion principle, each energy level can accommodate at most two electrons with opposite spins. Thus, the total energy of a free electron gas at $T = 0$ K must be finite. The Fermi-Dirac statistics implicitly contains the constraint imposed by the Pauli exclusion principle, as highlighted in Fig. 4.14.

The evaluation of the thermodynamical quantities implies calculations of integrals of the type

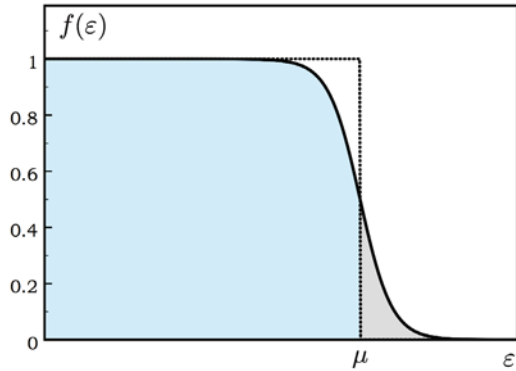
$$I = \int_{-\infty}^{\infty} H(\varepsilon) f(\varepsilon) d\varepsilon.$$

The integral form includes the case with $H(\varepsilon) = g(\varepsilon)$, so that the integral provides the electron density. This in turn requires the knowledge of the temperature and density dependence of chemical potential μ . While in Appendix 4.10 the calculation of integrals I is executed under more general conditions, the procedure is now outlined to determine density n , chemical potential μ and electronic energy per unit volume U , as functions of temperature.

Procedure

Step 1. Calculate

Fig. 4.14 Transport behavior of metals and Sommerfeld theory. The Fermi function at $T = 0$ K (dotted curve) and $T > 0$ K (solid curve). At finite temperature, the Pauli exclusion principle imposes that only electrons around $E = \mu$ can be thermally excited from electronic states below μ to electronic states above μ and represented by the gray region. This is a purely quantum effect and has crucial implications for the determination of the temperature dependence of specific heat and of conductivity in metals



$$n = \int_0^{\infty} g(\varepsilon) f(\varepsilon) d\varepsilon = C \int_0^{\infty} \frac{\sqrt{\varepsilon}}{e^{\beta(\varepsilon-\mu)} + 1} d\varepsilon, \quad (4.83)$$

with $C = (2m)^{3/2} / (2\pi^2 \hbar^3)$, integrating by parts and posing $x = \beta(\varepsilon - \mu)$. One has

$$n = -\frac{2}{3}C \int_{-\beta\mu}^{\infty} (x/\beta + \mu)^{3/2} \frac{d}{dx} \frac{1}{e^x + 1} dx. \quad (4.84)$$

Step 2. Use the conditions $\varepsilon_f \gg k_b T$, valid at room temperature, and $k_b T \ll \mu$ as in normal conditions, to extend the lower limit of integration to $-\infty$ and expand $(k_b T x + \mu)^{3/2}$ in power series. This results into:

$$n = -\frac{2}{3}C \int_{-\infty}^{+\infty} \left(\mu^{3/2} + \frac{3}{2}\mu^{1/2}k_b T x + \frac{3}{8}\mu^{-1/2}(k_b T)^2 x^2 \dots \right) \frac{d}{dx} \frac{1}{e^x + 1} dx.$$

Step 3. Exploit the parity properties of $h(x) = \frac{d}{dx} \frac{1}{e^x + 1} = h(-x)$, and the fact that

$$\int_{-\infty}^{\infty} x^2 \left(-\frac{d}{dx} \frac{1}{e^x + 1} \right) dx = \frac{\pi^2}{3}.$$

Step 4. Typically retain terms up to second order in $k_b T$, yielding the density n as:

$$n = \frac{2}{3}C\mu^{3/2} + \frac{\pi^2}{12}C \frac{(k_b T)^2}{\sqrt{\mu}}. \quad (4.85)$$

Step 5. Invert (4.85) to obtain the chemical potential $\mu(T)$ as a function of temperature. In particular, replace n with its expression in terms of ε_f , and $\mu(T)$ in the $(k_b T)^2$ term with its zero-temperature approximation $\mu(T \rightarrow 0) = \varepsilon_f$. One obtains to second order in $k_b T$:

$$\mu(T) = \varepsilon_f \left[1 - \frac{\pi^2}{12} \left(\frac{k_b T}{\varepsilon_f} \right)^2 \right]. \quad (4.86)$$

Step 6. Use the same procedure to calculate any other thermodynamic quantity, using the results of Appendix 4.10 for the suited I integral, along with the results (4.85) and (4.86) whenever useful.

Examples

Calculate the average electron energy per unit volume, as elaborated also in Problem 4.3. This is:

$$\begin{aligned} U(T) &= \int_0^\infty \varepsilon g(\varepsilon) f(\varepsilon) d\varepsilon = \frac{3}{5} n \varepsilon_f \left[1 + \frac{5}{12} \pi^2 \left(\frac{k_b T}{\varepsilon_f} \right)^2 \right] \\ &= U(0) + \frac{\pi^2}{6} g(\varepsilon_f) (k_b T)^2. \end{aligned} \quad (4.87)$$

The physics underlying all these many calculations is simple. First, the chemical potential is a function of temperature, the size of its deviation from ε_f being driven by the size of $k_b T$ on the scale of the Fermi energy ε_f . Considering the size of the latter in K degrees, significant deviations may occur at thousands or tens of thousands K. Second, as soon as temperature is switched on, the chemical potential lowers its value with respect to Fermi energy. Other significant deviations of μ from ε_f might occur under extreme conditions of reduced dimensions or strong correlations. Problem 4.4 develops the calculation of the chemical potential for 1D and 2D systems of non-interacting electrons. Third, the second line in (4.87) evidences that the average energy changes with respect to its value $U(0)$ at $T = 0$ by a quantity $\Delta U \simeq g(\varepsilon_f) (k_b T)^2$. This is the product of the average number of electrons $\sim g(\varepsilon_f) k_b T$ in an interval $k_b T$ around the Fermi energy, namely those which can be excited because of the Pauli principle, times the energy $k_b T$ carried on average by each of them.

More generally,

Concept

When Fermi-Dirac statistical behavior is used, thermodynamic properties show up as if they were contributed by only a reduced number of electron states $\sim g(\epsilon_f)k_bT$, namely those in the vicinity of the Fermi energy that can be excited to have an energy $\approx k_bT$. This is the essential concept brought in by the Sommerfeld theory of metals, which fixes many flaws of Drude model.

In Problem 4.5, the electronic current due to thermal effects is calculated for a metal, that is governed by the Richardson-Dushman law.

Quick Questions

Q24. Expression (4.87) states that the average energy per unit volume is quadratic in k_bT and depends on the electron density via the density of states at the Fermi energy. Is this the case in 2D?

Answer - No. While the reasoning would be similar, that is $U(T) \approx g(\epsilon_f)(k_bT)^2$ and thus the dependence on temperature would remain unchanged, in 2D the density of states is flat $g_{2D}(\epsilon) = m(\pi\hbar^2)^{-1}$ independent of the electron density.

4.3.5 Drude model vs. Sommerfeld theory

In the following the same physical quantities, that are relevant to transport behavior and have been calculated within the Drude model, are calculated within the Sommerfeld theory. As anticipated, many of the flaws encountered of Drude model are fixed when Sommerfeld theory is used instead. Of course, the knowledge of the temperature dependence of the scattering time still remains at a phenomenological level, this missing building block being tackled in Sec. 4.6.

4.3.5.1 Specific heat

The specific heat per particle is obtained by derivative of (4.87) with respect to temperature, obtaining the result:

$$c_V = \frac{\pi^2}{3n} g(\epsilon_f) k_b^2 T = \frac{\pi^2}{2} \frac{k_b T}{\epsilon_f} k_b = \frac{\pi^2}{2} \frac{T}{T_f} k_b. \quad (4.88)$$

This expression is valid in a wide range of temperatures. Relevant deviations of c_V from (4.88) are met at temperatures $T \simeq T_f = \epsilon_f/k_b \simeq 10^4$ K. The behavior of $c_V(T)$

as expressed by (4.88) agrees well with the experimental findings. The coefficient in front of $k_b T$ turns out to be consistent with experimental data if the effective electron mass is used to calculate the Fermi energy. Expression (4.88) solves one main failure of the classical free electron-gas model in predicting the room temperature specific heat, as discussed in Sec. 4.3.3.2. At any given temperature indeed, only a fraction T/T_f of electrons, those spanning an energy range $\simeq k_b T$ around the Fermi energy ϵ_f , can be excited to higher energy levels and thus contribute to the specific heat. At room temperature, $T/T_f \simeq 10^{-2}$, in agreement with the experimental findings.

Gathering these results together with the vibrational contribution determined in Chap. 3, it is found that while $T < \Theta_D$ the total specific heat per particle is given by

$$c_V = \frac{12\pi^4}{5} k_b \left(\frac{T}{\Theta_D} \right)^3 + \frac{\pi^2}{2} k_b \frac{k_b T}{\epsilon_f}. \quad (4.89)$$

It is convenient to display this relation as a c_V/T vs. T^2 function, so that the angular coefficient of the resulting straight line represents the phonon-related T^3 coefficient of the vibrational specific heat, whereas the intercept represents the electron-related T -coefficient. Fig. 4.15 displays the measured specific heat for two different metals, potassium and copper, evidencing the excellent agreement with the theory (4.89).

The ratio between electron and phonon contributions to specific heat is a relevant quantity as well, to predict the temperature range where one prevails against the other. It is easily seen that the vibrational contribution to specific heat becomes predominant at temperatures larger than

$$T_0 = \left(\frac{5}{24\pi^2} Z \frac{\Theta_D}{T_f} \right)^{\frac{1}{2}}, \quad (4.90)$$

that is of the order of a few K, Z being the number of valence electrons. In Fig. 4.16 the specific heat of silicon doped with phosphorous is displayed as a function of temperature, evidencing an excellent agreement with the theoretical prediction of an electron-dominated contribution at lower temperatures and a vibrational-dominated one at higher temperatures.

4.3.5.2 Vibrational energy and specific heat of nanostructures

The question arises how the results obtained for the electronic contribution to specific heat would change in the case of nanostructures. This question is answered at the phenomenological level here, following the same reasoning lines as in Chap. 3 for the phonon contribution. There, normal variables had been conveniently introduced to treat the vibrational modes of both infinite and nanometric crystals, which satisfy the equations of motion of the harmonic oscillator with the suited boundary conditions. On this basis, the thermodynamics of vibrational modes results to be the same for both macroscopic and nanometric systems.

Fig. 4.15 Transport behavior of metals and Sommerfeld theory. Specific heat of potassium and copper at low temperatures. For easier comparison with (4.89), the quantity c_V/T vs. T^2 is displayed, evidencing the excellent agreement with the theoretical predictions. Symbols: experimental data. Solid lines: linear fit to the data as in (4.89) [15, 16]

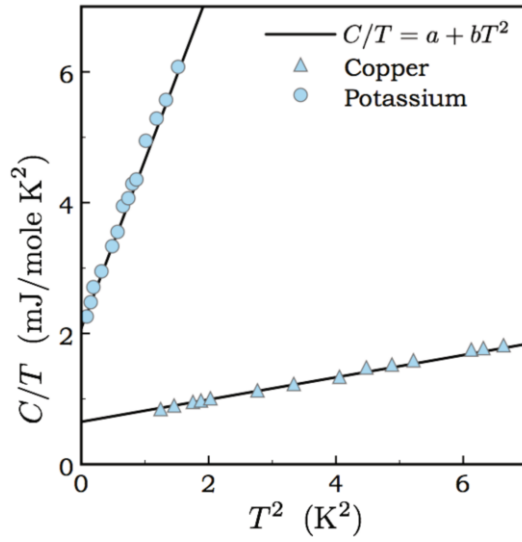
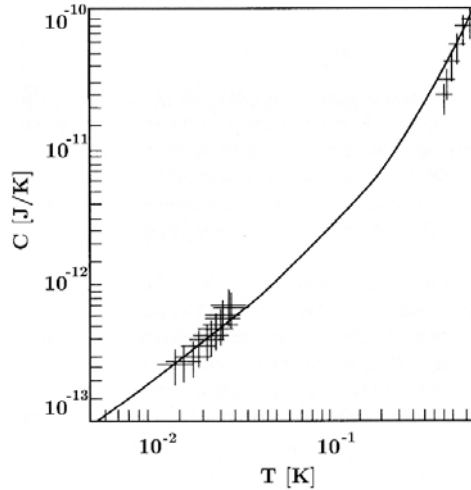


Fig. 4.16 Transport behavior of metals and Sommerfeld theory. Specific heat of silicon doped with phosphorous at very low temperatures, showing that the electron contribution dominates at lower temperatures and the vibrational one at higher temperatures. Crosses: experimental data with related errors. Solid curve: fit to the data according to the theoretical prediction [17]



Once the theoretical connections between the two systems is established, the next step though is to notice that measurements of total energy or specific heat in nanostructures are performed on ensembles of similar but not identical nanostructures. Therefore, the experimental data refer to ensembles of objects. For this reason, the expression for the specific heat has to account for all the possibilities which might come about. Within a phenomenological discussion, at low temperatures one may write:

$$C_{np} = \gamma_n T + A_n T^2 + B_n T^3 + \frac{A_0}{T^2} e^{(A - \frac{B}{T})}, \quad (4.91)$$

where the meaning of the different terms is as follows. The term $\gamma_n T$ represents the contribution of free electrons, along the derivations developed so far for electrons in infinite crystals. The second and third terms refer to Debye-like vibrational contributions in 2D and 3D geometries discussed in Chap. 3. The last term comes from the Einstein-like contribution to specific heat in the $k_b T \ll \hbar \omega_0$ limit discussed in Chap. 3.

The new key point here is that the measured coefficient γ_n is smaller than the value predicted within the Sommerfeld theory of metals and it depends on the structure size. For a metallic system with volume V , γ_n has indeed been found to be:

$$\gamma_n = \frac{\pi^2}{2} k_b^2 \frac{N}{\varepsilon_f} = 2\pi^2 k_b^2 \frac{1}{3\delta}, \quad (4.92)$$

with $\delta \equiv 4\varepsilon_f/(3N)$. If the same definition of δ were used for a nanometric system as well, one would obtain:

$$\gamma_n = \frac{\pi^2}{2} \frac{1}{(3\pi^2)^{\frac{2}{3}}} N^{\frac{1}{3}} V^{\frac{2}{3}}, \quad (4.93)$$

where N represents the number of conduction electrons. Two parameters may lead to smaller values of γ_n : lower values of N due to the presence of interfaces where electrons might localize with partial oxidation, and smaller nanostructure volumes V .

Fig. 4.17 displays the measured specific heat of iron-based nanostructures. The solid line represents a fit to expression (4.91), showing an excellent agreement [18].

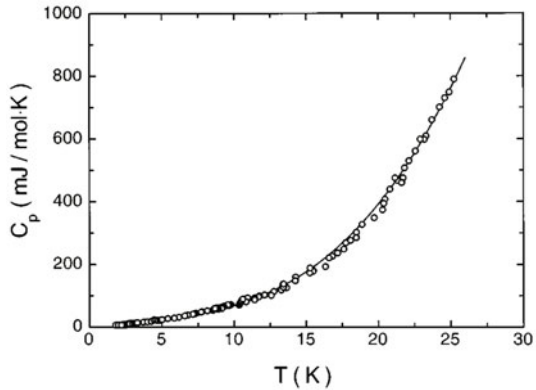


Fig. 4.17 Transport behavior of metals and Sommerfeld theory: application to nanostructures. Specific heat of Fe-based nanostructures at low temperatures, evidencing the electronic contribution. Symbols: experimental data. Solid curve: fit to the theoretical prediction (4.91) [18]

4.3.5.3 Thermal conductivity and thermoelectric coefficient

Expressions (4.66) remain formally valid, but of course the dependence of v^2 and c_V on T changes for similar reasons as those discussed at the end of Sec. 4.3.5.1. Indeed, with the definition $v_f = \hbar k_f / m$, one has $v^2 \simeq v_f^2$ consistently with the concept that at finite temperature T only a small fraction $g(\epsilon_f)k_b T / \epsilon_f$ of electrons in the vicinity of the Fermi level is excited. Besides, as already seen $c_V \propto T$. Overall then, $\chi_c(T) \propto T\tau(T)$ and $\chi_t(T) \propto T$. While definite conclusions cannot be inferred yet on the thermal conductivity until the function $\tau(T)$ is unknown, the linear dependence of the thermoelectric coefficient on T is in agreement with experimental data.

4.3.5.4 Electrical conductivity

Once again, expression (4.50) remains formally valid, though the dependence $\tau(T)$ is still missing. Nevertheless, a qualitatively different picture emerges. The ground state of the quantum free electron gas at $T = 0$ is characterized by having all the electronic levels occupied up to the Fermi level k_f . This can be represented in \mathbf{k} space, by stating that all the occupied states are those whose wavevector lies inside the Fermi sphere, i.e. $|\mathbf{k}| \leq k_f$. Consider the total wavevector $\mathbf{k}_{\text{tot}} = \sum_{i=1}^N \mathbf{k}_i$, obtained by summing up the wavevectors of all the electrons. One has $\mathbf{k}_{\text{tot}} = 0$, because for each occupied electronic state with wavevector \mathbf{k} there is an occupied electronic state with wavevector $-\mathbf{k}$.

In the presence of an external electric field \mathbf{E} , the wavevector of each electron changes by $\delta\mathbf{k} = -e\mathbf{E}\delta t/\hbar$ in the time interval δt . This corresponds to a rigid translation of the Fermi sphere by $\delta\mathbf{k} \ll k_f$ and to a total momentum change $N\delta\mathbf{k}$. To have an idea about typical values, copper conductivity is $\sigma_{\text{Cu}} \simeq 6 \cdot 10^5 \Omega^{-1} \text{cm}^{-1}$ and the measured electron density $n_{\text{Cu}} \simeq 8.5 \cdot 10^{22} \text{cm}^{-3}$. From (4.50), $\tau \simeq 2.5 \cdot 10^{-14} \text{s}$. Under an applied electric field with modulus $E \simeq 1 \text{V/cm}$, the displacement of the Fermi sphere is $eE\tau/\hbar \simeq 3.8 \cdot 10 \text{cm}^{-1} \ll k_f \simeq 1.35 \cdot 10^8 \text{cm}^{-1}$. A schematic illustration of such a process is pictured in Fig. 4.18. States which were occupied (empty) in the ground state, become empty (filled) after the application of the electric field. The rigid translation of the Fermi sphere would last indefinitely, if no scattering event took place. However, the interaction of the electrons with the lattice and/or with impurities tends to restore the occupation of the ground state. This induces electronic transitions from the occupied states marked with gray dots in Fig. 4.18, to the empty ones marked with white dots.

In essence, in the classical Drude model an equilibrium, stationary condition sets in the interval between the action of the electric field and that of the scattering mechanisms, where the Fermi sphere keeps being translated by $-e\mathbf{E}\tau/\hbar$.

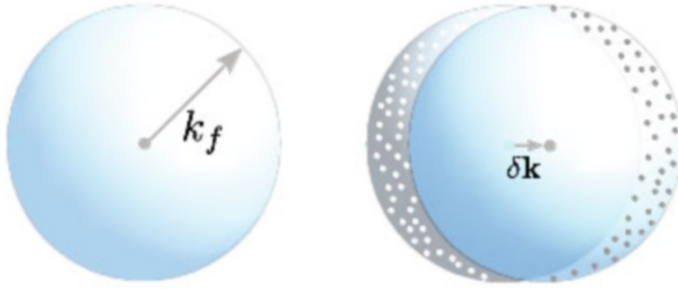


Fig. 4.18 Quantum model for the transport in a free electron gas. Left panel: the Fermi sphere representing the ground state. Right panel: Fermi sphere in the presence of an applied electric field, resulting into its rigid translation. As a consequence, electronic states that were occupied in the ground state become empty, and viceversa. Gray (white) dots: occupied (empty) states

4.3.5.5 Mean free path

A completely different picture arises with respect to the Drude model. Indeed, according to the Pauli's exclusion principle, only electrons within about $k_b T$ around the Fermi energy can be scattered. Therefore, the average velocity of the scattered electrons, responsible for the resetting of the acquired momentum, is the Fermi velocity v_f rather than the classical velocity. Since $v_f \sim 10^8$ cm/s, the mean-free path $\ell = v_f \tau$ gets as large as $\sim 10^{-6}$ cm, much longer than the distance between neighbor ions in their equilibrium lattice positions. This finding supports the reasoning according to which the fixed ions cannot be responsible for the scattering mechanism. Besides, a so long mean-free path strengthens the validity of the semiclassical treatment: the linear size Δr of a wavepacket may more easily satisfy the condition $\Delta r \ll \ell$. Finally, since v_f is essentially independent of temperature, the relaxation time $\tau = \ell / v_f$ gets the same temperature dependence of the mean free path. In Sec. 4.6.2 is discussed how it occurs that the mean free path be inversely proportional to temperature for values of T around room temperature. To conclude, the evaluations performed in Sec. 4.3.3.1 are modified as follows: $\sigma(T = 273 \text{ K}) / \sigma(T = 77 \text{ K}) \simeq \tau(T = 273 \text{ K}) / \tau(T = 77 \text{ K}) = (77/273) \simeq 0.28$. The experimental data represented in Fig. 4.12 are closer to this estimate rather than to that calculated in Sec. 4.3.3.1.

4.3.5.6 Wiedemann and Franz law

Considering that $v^2 \simeq v_f^2$ and using the Sommerfeld expression for c_V , one finds that

$$\frac{\chi_c}{\sigma T} = \frac{\pi^2}{3} \left(\frac{k_b}{e} \right)^2 = 2.44 \cdot 10^{-8} \frac{\text{W}\Omega}{\text{K}^2}. \quad (4.94)$$

Thus, Sommerfeld theory predicts the same result as the Drude model: the ratio $\chi_c/\sigma T$ is the same for all the materials, it being independent of τ . This result agrees with experimental findings. No news are brought by the Sommerfeld theory about the reasons for the different behavior at low temperatures $T \ll \Theta_D$.

4.3.5.7 Magnetic susceptibility

The magnetic susceptibility has been calculated in Sec. 2.7.3 within the classical theory. In the Sommerfeld theory, this can be qualitatively estimated by considering that in a wide range of temperatures and magnetic fields, the conditions $\mu_s B \ll k_b T \ll \epsilon_f$ occur, where $\mu_s = 2\mu_b$. Then, as discussed above, one notices that the fraction of electrons that can be excited at finite temperature is $\sim g(\epsilon_f)k_b T$. Finally, among these electrons participating to the dynamics, the ratio of those with spin up to those with spin down is of the order of their energy separation $\mu_s B$ on the scale of $k_b T$, that is $\sim \mu_s B/(k_b T)$. Collecting all these estimates together, one finds that magnetization and magnetic susceptibility are given by

$$\begin{aligned} M &\sim \mu_s g(\epsilon_f) k_b T \frac{\mu_s B}{k_b T} \\ \chi_m &\sim \mu_s^2 g(\epsilon_f) = \frac{3}{2} \frac{n \mu_s^2}{k_b T_f}, \end{aligned} \quad (4.95)$$

independently of temperature.

Reverting from the estimate to a detailed calculation, the spin up n_p (parallel to the field) and spin down n_{ap} (antiparallel to the field) electron densities are first conveniently written:

$$\begin{aligned} n_p &= \frac{1}{2} \int_0^\infty g(\epsilon) f(\epsilon - \mu_s B) d\epsilon \\ n_{ap} &= \frac{1}{2} \int_0^\infty g(\epsilon) f(\epsilon + \mu_s B) d\epsilon, \end{aligned} \quad (4.96)$$

where up and down obviously refer to the direction of the applied magnetic field. The integrals are of the form calculated in Appendix 4.10. Following the procedure outlined in the Appendix, one finds:

$$\begin{aligned} n_p &= \frac{1}{2} \int_0^{\mu+\mu_s B} d\epsilon g(\epsilon) + \frac{\pi^2}{6} (k_b T)^2 \frac{1}{2} \frac{dg(\epsilon)}{d\epsilon} \Big|_{\epsilon=\mu+\mu_s B} \\ n_{ap} &= \frac{1}{2} \int_0^{\mu-\mu_s B} d\epsilon g(\epsilon) + \frac{\pi^2}{6} (k_b T)^2 \frac{1}{2} \frac{dg(\epsilon)}{d\epsilon} \Big|_{\epsilon=\mu-\mu_s B}. \end{aligned} \quad (4.97)$$

Using the explicit expression of $g(\epsilon)$ along with the condition $\mu_s B \ll \mu$, one has:

$$n_p - n_{ap} = \frac{1}{2} \int_{\mu-\mu_s B}^{\mu+\mu_s B} g(\epsilon) d\epsilon + \frac{1}{2} \frac{\pi^2}{6} (k_b T)^2 \int_{\mu-\mu_s B}^{\mu+\mu_s B} \frac{d^2 g(\epsilon)}{d\epsilon^2} d\epsilon$$

$$\begin{aligned}
&\simeq \mu_s B g(\mu) + \frac{\pi^2}{6} (k_b T)^2 \mu_s B \frac{d^2 g(\epsilon)}{d\epsilon^2} \Big|_{\epsilon=\mu} \\
&= \mu_s B g(\mu) \left[1 - \frac{\pi^2}{24} \left(\frac{k_b T}{\mu} \right)^2 \right], \tag{4.98}
\end{aligned}$$

Further simplifications arise by inserting the energy expansion of $g(\mu)$:

$$g(\mu) = g(\epsilon_f) + (\mu - \epsilon_f) g'(\epsilon_f) = g(\epsilon_f) \left[1 - \frac{\pi^2}{24} \left(\frac{k_b T}{\epsilon_f} \right)^2 \right].$$

Thus, one finds:

$$n_p - n_{ap} = \mu_b B g(\epsilon_f) \left[1 - \frac{\pi^2}{12} \left(\frac{k_b T}{\epsilon_f} \right)^2 \right]. \tag{4.99}$$

Finally, the susceptibility is given by the expression

$$\chi_m = \frac{3}{2} \frac{n \mu_b^2}{k_b T_f} \left[1 - \frac{\pi^2}{12} \left(\frac{k_b T}{\epsilon_f} \right)^2 \right], \tag{4.100}$$

essentially temperature independent and smaller than that predicted by the Drude model in (4.82) at room temperature. In conclusion, Sommerfeld theory repairs the flaws of the Drude model in predicting the experimental behavior of χ_m .

If diamagnetic effects are taken into account [3], (4.100) becomes:

$$\chi_m = \frac{7}{6} \frac{\mu_b^2 n}{k_b T_f} \left[1 - \frac{3}{28} \left(\frac{k_b T}{\epsilon_f} \right)^2 \right].$$

4.3.5.8 Classical Hall effect

Expression (4.81) remains valid in Sommerfeld theory as well. Unfortunately, this time also the inconsistency of the theoretical prediction with the experimental findings persists. To see this, consider the quantity $U = n_0/n \equiv -1/(R_H n e c)$ for strong magnetic fields B and very low temperatures T , with R_H the measured Hall coefficient and n the density as it is barely estimated from the material chemical valence. In practice, n_0 represents the density for which the Drude and Sommerfeld forms of the Hall coefficient $-1/(n e c)$ agree with the experimental data: if Drude and Sommerfeld expressions were in agreement with the data, one should have $U = 1$. However, Table 4.1 evidently shows that this is not the case. While for selected alkali metals U is close to the theoretical value 1, quite significant deviations are found for metals in the second column of the periodic table and for metals with more complex electronic structures such as Cu, Ag, and Au. The value of U can as well become negative, indicating a different sign in the prediction $-1/(n e c)$. An example over all

strikingly illustrates this point and is displayed in Fig. 4.19. The quantity U for Al is plotted as a function of $\omega_c \tau$, as measured in experiments [19]. It is seen that n_0/n depends on the magnetic field size, saturating to the value $-1/3$ while B increases. In essence, the metal behaves as if n_0 represented a unit positive charge, the chemical valence of Al being 3. Or else, as if the ensemble of three valence electrons of Al behave as one hole.

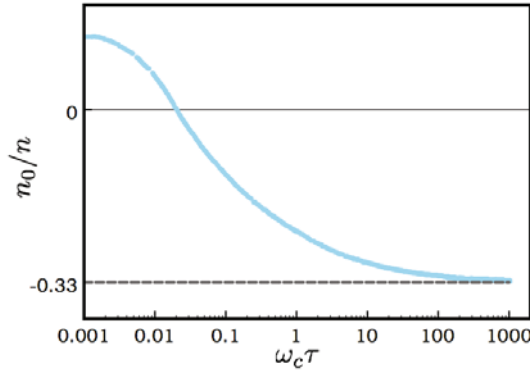


Fig. 4.19 Transport behavior of metals and Sommerfeld theory. Classical quantum Hall effect. Deviations from Drude and Sommerfeld prediction. The quantity $U = n_0/n \equiv -1/(R_H n e c)$ vs. $\omega_c \tau$ as measured in Al (see text), with R_H the measured Hall coefficient. The Drude and Sommerfeld prediction state that $U = 1$. Significant deviations show up instead with varying B in $\omega_c \tau$, up to the peculiar value $-1/3$ for $\omega_c \tau \gg 1$, as if a positive - instead than negative- charge were involved in the Hall process, with size equal to $1/3$ of the valence of Al [19]

4.3.5.9 Quantum Hall effect

Introduction of a quantum statistical treatment of electrons behavior as in Sommerfeld theory calls for a reconsideration of the Hall effect. Indeed, when the quantum nature of electrons is taken into account, the classical orbital motion that the particles are seen to travel in the presence of a magnetic field turns into quantized energy states, named Landau levels. These result from the solution of the Schrödinger equation for electrons in a uniform magnetic field, and are calculated in Appendix 4.11 along with related eigenenergies ε and density of states $G(\varepsilon)$. The energies of the Landau levels turn out to be $\varepsilon = \hbar^2/(2m)k_z^2 + (p + \frac{1}{2})\hbar\omega_c$ in terms of the integer p . They are depicted in Fig. 4.20, where the parabolic dispersion characteristic of single-particle orbitals on top of the discrete levels is evident. The density of states for given n is expressed as $G(\varepsilon) = eB\sqrt{2m}[8\pi^2\hbar^3 c\sqrt{\varepsilon - \varepsilon_n}]^{-1}$, with $\varepsilon_n \equiv (n + 1/2)\hbar\omega_c$. This is displayed in Fig. 4.21.

The existence of quantized energy states qualitatively modifies the physics of crystal electrons in the presence of magnetic fields. Even for weak magnetic fields, oscillations are predicted to occur on a faster scale than the cyclotron period as dictated by the existence of the quantization conditions and by the interplay between

Fig. 4.20 Transport behavior in systems with reduced dimensionality and in the presence of magnetic fields. Landau levels. Energy bands of a free electron in a magnetic field, evidencing the parabolic dispersion on top of the discrete levels

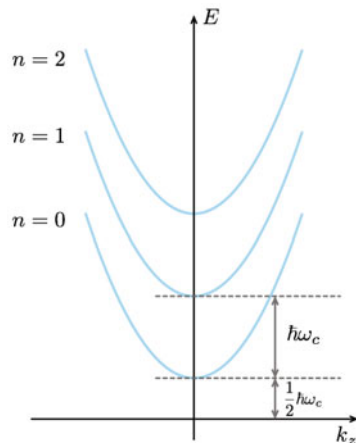
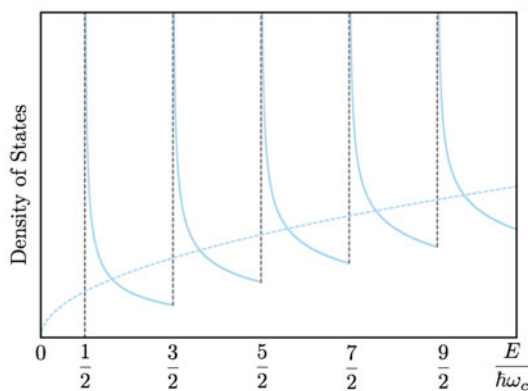


Fig. 4.21 Transport behavior in systems with reduced dimensionality and in the presence of magnetic fields. Landau levels. The density of states for a free electron in the presence of a magnetic field. For comparison, the density of states for a free particle in the absence of a magnetic field is shown, represented by the dashed curve



Fermi level and Landau levels. For strong magnetic fields, significant qualitative changes apply to Hall effect manifestations and Quantum Hall Effect occurs. Quantum Hall Effect has significant applications in precision measurements and in the design of nanodevices. For the Quantum Hall effect to be visible, extreme conditions of quantum degeneracy are required, such as those that can be obtained in structures where the particles de Broglie wavelength be comparable to the system size. Now, the larger is the de Broglie wavelength λ_{dB} , the smaller is the kinetic energy $\approx \hbar^2/(2m\lambda_{dB}^2)$. As shown later on in this Chapter, reduced electron densities may do the job of increasing the interaction-to-kinetic energy ratio. Also, reduction of the system size along one or more directions corresponds to a reduction of effective system dimensionality. Thus, extreme conditions of quantum degeneracy can be met either when interaction energies are enhanced with respect to the kinetic part, or when the system dimensionality is reduced, or both.

Along with the concept underlying the Hall effect at hand, two dimensions are in fact needed. Consider thus an interface structure where the following conditions can

be set: the motion of electrons be free along the surface and confined in direction orthogonal to it. Then apply an uniform magnetic field along this latter direction: quantization of the electron energy levels into the Landau levels occurs, with a high degeneracy favored by the reduced system dimensionality. In particular, when the condition $\hbar\omega_c \gg k_b T$ of strong magnetic fields and low temperatures is conveniently met, one can safely consider that the discrete Landau levels be either completely full or completely empty. Besides, scattering with phonons is highly suppressed. In essence, under these conditions the physical concept underlying Hall effect in stationary conditions remains valid as described by (4.80), except for two crucial differences. The first is that now n represents the surface electron density, say n_s . Thus, the potential difference measured along the y direction is

$$U_y = -E_y y = \frac{B}{n_s e c} i_x \quad (4.101)$$

with i_x the current intensity along the x direction. From now on,

Definition

$U_y \equiv U_H$ is referred as Hall potential.

Also,

Definition

The quantity

$$\rho_H = \frac{B}{n_s e c} = \frac{U_H}{i_x} \quad (4.102)$$

is called Hall resistance.

The second difference is conceptual. In the considered case, two next levels are separated by the energy $\hbar\omega_c$ and each level is characterized by degeneracy

$$d = \frac{e}{2\pi c \hbar} B,$$

as calculated in Appendix 4.11. If an integer number λ of levels is fully occupied, then the surface density $n_s = \lambda d$ is a multiple of d . Substitution of n_s in (4.102) yields

$$\rho_H = \frac{2\pi\hbar}{e^2\lambda}. \quad (4.103)$$

Significant consequences follow from the two reported differences with respect to the classical Hall effect. First, the Hall resistance U_H/i_x is predicted to be quantized, that is composed of a set of plateaus with values that are submultiples of the basic unit h/e^2 , built up from the fundamental physical constants h and e . Second,

in correspondence with this full filling of the 2D Landau levels, scattering mechanisms are suppressed because vacant states are unavailable which electrons may scatter into. Thus, when the levels are full filled, one expects infinite conductivity (zero resistance) and therefore vanishing voltage drop U_p along the x direction. The above predictions are clearly observed in the experimental data reported in Fig. 4.22 from [23]. The plateaus of U_H/i_x are clearly visible at the predicted values, with ordinate axis on the left. In correspondence of them, the voltage drop along the direction of the current is seen to collapse to zero, with ordinate axis on the right. The visibility of the effect becomes as much neater as much stronger is the magnetic field: stronger magnetic fields lead to higher degeneracy conditions and wider plateaus.

One can also imagine to lower the temperatures and increase the value of the magnetic field until only the lowest Landau level is occupied and only part of it. Under such especially extreme quantum-degeneracy conditions, the above treatment breaks down. In fact, experiments performed under these conditions show that the Hall resistance is quantized with fractional values of λ , the actual value depending on the filling of the lowest Landau level. This is referred to as Fractional Quantum Hall Effect (FQHE) in comparison with the Integer QHE discussed so far. Integer QHE can be understood in terms of Landau levels and single-particle orbitals in the presence of a magnetic field within reduced dimensions. FQHE instead requires some more elaborations and is more often explained in terms of single-particle orbitals of effective particles, composed of electrons bound to an even number of vortices: these quasiparticles turn out to have anyon, that is neither boson nor fermion, fractional statistics. This is the essential content of the so-called composite-fermion theory, describing the very special collective character of this new quantum liquid and new quantum state of matter. In addition to magnetically ordered materials and superconductors or superfluids described in Chap. 2, FQHE states represent other quite interesting quantum ground and excited states of matter that can be conceived and realized.

Notice that IQHE and FQHE can be exploited for quantum transport applications in miniaturized devices, or else to get extremely precise measurements of h/e^2 , actually up the level of a few ppb, and provide accurate standards for the measurement of, e.g., resistances. In conclusion,

Concept

Extreme conditions of quantum degeneracy, such as can be met with reduced dimensions, low temperatures and strong external magnetic fields, favor the realization of new quantum states due to the occurrence of Integer and Fractional Quantum Hall Effect. In Quantum Hall Effects, very peculiar quantum transport behavior emerged, such as quantized resistance and infinite conductivity, and can be used to perform precision measurements of fundamental constants.

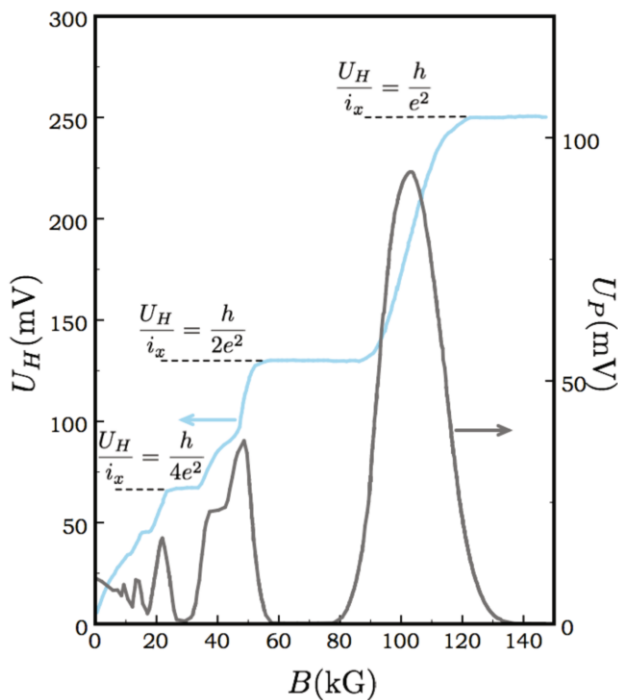


Fig. 4.22 Transport behavior of metals. Quantum Hall Effect. Measured Hall potential U_H/i_x (left) and voltage drop U_p in the direction of the current flow, as functions of the magnetic-field strength B . The increase of B leads to full filling of progressively lower Landau levels. Under the experimental conditions of low temperatures and strong magnetic fields, the Landau levels are essentially either full or empty. The quantized Hall resistance is visible as the plateaus in U_H/i_x at full level filling with $\lambda = 1, 2, 4, \dots$. The infinite conductivity is inferred by the voltage drop U_p , collapsing to zero in correspondence of the plateaus [23]

Quick Questions

Q25. A material, considered to have a 2D character, undergoes of phase transition to a new ground state. Before the transition, the Hall resistance is measured to be ρ_H . After the transition, it is measured to be $\rho_H/4$. Assume that the Hall magnetic field does not affect the new ground state nor the 2D system character, and that in both cases the same number of levels has been occupied. Is there an hypothesis that can be inferred on the nature of the new ground state?

Answer - If everything else (\hbar and λ in $\rho_H = 2\pi\hbar(e^2\lambda)^{-1}$) remains unchanged as assumed in the text, the only other possibility is that in the new ground state the carriers behave as if they had twice the charge of an electron, thus as if the carriers were pairs of electrons. Notice that such a situation can be realized in experiments.

4.4 Semiconductors

Semiconductors have been characterized in Chap. 2 with respect to their electronic structure. It has been shown that at zero temperature the valence bands are full, the conduction bands empty, and that a gap manifests between the top of the valence band and the bottom of the conduction band where electronic states cannot exist. It is well known that

Properties

P1. Semiconductors are characterized by a finite electric conductivity at finite temperatures, its size depending on the number of impurities in the material, either intrinsic or introduced by doping procedures. At higher temperatures the conductivity drops to a smaller value that is independent of the number of impurities. Finally, vanishing temperatures correspond to vanishing conductivities.

In contrast, conductivity in metals with impurities increases while the temperature drops: in particular, it reaches a value that at zero temperature depends on the doping level, as seen in Fig. 4.8. Besides the temperature dependence, the rate of change of the conductivity after variations of the doping level is different, very large in semiconductors and small in metals. In practice, in semiconductors doping levels of the order of 10^{14} atoms per cm^3 already determine quite significant and measurable changes in the conductivity behavior, even though the number of impurities is nine orders of magnitude smaller than the Avogadro number. This peculiar property has interesting and important applications. Fig. 4.23 shows an example of such a behavior, displaying the resistivity of Sb-doped Ge as a function of inverse temperature for a number of doping concentrations. The study of transport in semiconductors requires the introduction of a number of useful intermediate concepts. This is the aim of the following sections.

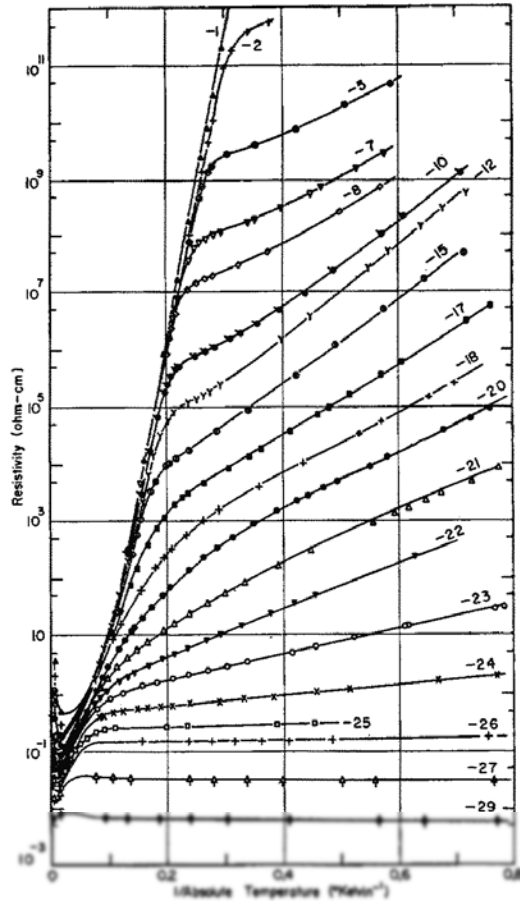


Fig. 4.23 Transport behavior of semiconductors. Experimental data on resistivity of Ge doped with Sb as a function of T^{-1} (K^{-1}) for a number of doping concentrations and numbered as in the legend. From 1 to 29 the concentrations are, in cm^{-3} : $5.3 \cdot 10^{14}$, $9.3 \cdot 10^{14}$, $1.6 \cdot 10^{15}$, $2.3 \cdot 10^{15}$, $3.0 \cdot 10^{15}$, $5.2 \cdot 10^{15}$, $8.5 \cdot 10^{15}$, $1.3 \cdot 10^{16}$, $2.4 \cdot 10^{16}$, $3.5 \cdot 10^{16}$, $4.5 \cdot 10^{16}$, $5.5 \cdot 10^{16}$, $6.4 \cdot 10^{16}$, $7.4 \cdot 10^{16}$, $8.4 \cdot 10^{16}$, $1.2 \cdot 10^{17}$, $1.3 \cdot 10^{17}$, $2.7 \cdot 10^{17}$, $9.5 \cdot 10^{17}$ [20]

4.4.1 Chemical potential in intrinsic and doped semiconductors

The introduction of doping impurities affects first of all the chemical potential. To understand how this works, a simplest but physically meaningful model is set up to schematize a semiconductor. This is one single valence and one single conduction band separated by an energy gap E_g . Real semiconductors often do not fit this scheme, since the conduction band bottom and the valence band top can occur in correspondence of different points of the BZ, or else because band degeneracy may occur. The latter is the case of the valence band of the most used Si. Nevertheless, the simplest model works to establish the basic concepts.

In the model, electron density n_c in the conduction band and hole density p_v in the valence band can be expressed in the form

$$n_c = \int_{\epsilon_c}^{\epsilon_c + \Delta_c} d\epsilon g_c(\epsilon) f(\epsilon)$$

$$p_V = \int_{\epsilon_V - \Delta_V}^{\epsilon_V} d\epsilon g_V(\epsilon)(1 - f(\epsilon)), \quad (4.104)$$

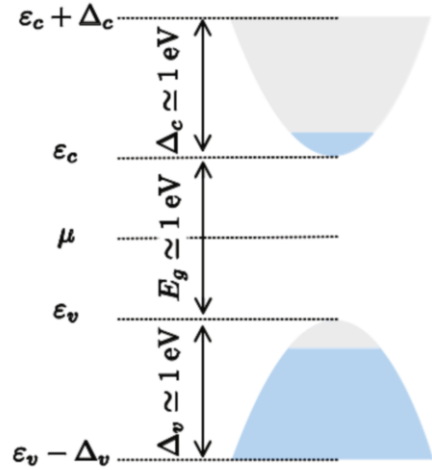
in terms of the energy of the conduction-band bottom ϵ_c and of the valence-band top ϵ_V . The corresponding widths Δ_c and Δ_V of the two bands, and the density of states $g_c(\epsilon)$ and $g_V(\epsilon)$, are sketched in [Fig. 4.24](#). Notice that, while each energy level in the conduction band contributes to the electron density with a weight given by the probability $f(\epsilon)$ that it is occupied, in the valence band, where the hole density is being computed, the weight is given by the probability that the energy level is not occupied, that is by $1 - f(\epsilon)$. The numbers of both electrons and holes depend on the chemical potential through the Fermi distribution function. In the case of metals, it has been found that the chemical potential μ at zero temperature coincides with the Fermi energy, that is the energy of the highest occupied state. At $T > 0$ instead, μ slowly changes with increasing temperature, in particular it decreases in 3D and increases in 1D.

The position of the chemical potential in semiconductor materials is more tricky, because of the presence of the gap. At $T = 0$, the valence-band states are fully occupied and the conduction-band states fully empty: thus, the chemical potential value must be higher than ϵ_V and lower than ϵ_c , for the Fermi distribution function be 1 for energies of the valence states and 0 for the conduction states. When doping is introduced, it is useful to anticipate the basic results detailed later on this Chapter. In essence, electronic states can appear that are localized with energies within the gap, and that can be either occupied or empty. If occupied, the chemical potential gets a value that is above the highest among the energies of the occupied states. If empty, the chemical potential lies in between the top of the valence band and the lowest among the energies of the non-occupied states.

Now, semiconductors are characterized by gap energies of the order of $E_g \sim 1$ eV and by a very small number of free electrons on the scale of the Avogadro number, though still sufficiently large to determine measurable electric currents that are activated by electron thermal energies $\sim k_b T$. In order to make all these facts evident and to simplify the calculations, a few model assumptions are adopted in the following treatment. First, μ is taken to be inside the gap, with $(\epsilon_c - \mu)/(k_b T) \gg 1$ and $(\mu - \epsilon_V)/(k_b T) \gg 1$ at room temperature. This latter assumption will be checked a posteriori after the integrals (4.104) have been executed. Second, since the conditions $\Delta_c/(k_b T) \gg 1$ and $\Delta_V/(k_b T) \gg 1$ hold also well above the room temperature, the contribution from the lowest states in the valence band and from the highest states in the conduction band can be neglected. The schematic view of the different energy levels and of the relationship among them is depicted in [Fig. 4.24](#). Therefore, n_c and p_V are given by the simplified expressions:

$$\begin{aligned} n_c &= \int_{\epsilon_c}^{\infty} g_c(\epsilon) f(\epsilon) d\epsilon \\ p_V &= \int_{-\infty}^{\epsilon_V} g_V(\epsilon) (1 - f(\epsilon)) d\epsilon. \end{aligned} \quad (4.105)$$

Fig. 4.24 Transport behavior in semiconductors. A schematic band structure along with the relevant energy levels and widths (see text): valence-band top ε_v and width Δ_v ; conduction-band bottom ε_c and width Δ_c ; chemical potential μ , as intended for an intrinsic semiconductor. In typical semiconductors, energy gaps $E_g = \varepsilon_c - \varepsilon_v$ are of the order of 1 up to a few eV and bandwidths of the order of 1 eV



If μ lies in the gap, the number of conduction electrons and of valence holes is usually much smaller than the Avogadro number. If doping is heavy instead, say $> 10^{20} \text{ cm}^{-3}$ or in the presence of strong external fields, the number of conduction electrons or valence holes can be so large that either $\mu > \varepsilon_c$ or $\mu < \varepsilon_v$. The latter cases excluded, only the states with lower energy in the conduction band are occupied and only the states with higher energy in the valence band are empty. Therefore, with reference to Chap. 3, the parabolic approximation of the two bands around the respective edges can be assumed. Under these conditions, one finds the densities of states:

$$\begin{aligned} g_c(\varepsilon) &= \frac{1}{2\pi^2} \left(\frac{2m_c^*}{\hbar^2} \right)^{3/2} (\varepsilon - \varepsilon_c)^{1/2} \\ g_v(\varepsilon) &= \frac{1}{2\pi^2} \left(\frac{2m_v^*}{\hbar^2} \right)^{3/2} (\varepsilon_v - \varepsilon)^{1/2}. \end{aligned} \quad (4.106)$$

Here, $m_c^* = (m_1 m_2 m_3)^{1/3}$ and $m_v^* = |m_1 m_2 m_3|^{1/3}$ are the average effective masses at the conduction bottom and valence top, respectively, with m_i the corresponding principal values of the mass tensor. One therefore obtains:

$$\begin{aligned} n_c &= \frac{1}{2\pi^2} \left(\frac{2m_c^*}{\hbar^2} \right)^{3/2} \int_{\varepsilon_c}^{\infty} (\varepsilon - \varepsilon_c)^{1/2} \frac{1}{1 + e^{\frac{\varepsilon - \mu}{k_b T}}} d\varepsilon \\ &= \frac{1}{2\pi^2} \left(\frac{2m_c^* k_b T}{\hbar^2} \right)^{3/2} F_{1/2} \left(\frac{\mu - \varepsilon_c}{k_b T} \right). \end{aligned} \quad (4.107)$$

The following quantities are conveniently introduced: $N_0(T) \equiv (2\pi^2)^{-1} (2m_c^* k_b T / \hbar^2)^{3/2} \simeq 3.3 \cdot 10^{18} (\bar{m}_c T / 100)^{3/2} \text{ cm}^{-3}$, with \bar{m}_c the effective mass in units of the

free-electron mass, and

$$F_{1/2}\left(\frac{\mu - \varepsilon_c}{k_b T}\right) = \int_{-\frac{\mu - \varepsilon_c}{k_b T}}^{\infty} dx \left(\frac{\mu - \varepsilon_c}{k_b T} + x\right)^{1/2} \frac{1}{1 + e^x}. \quad (4.108)$$

$F_{1/2}(x)$ is the Dirac integral $F_j(x)$ with $j = 1/2$, its properties being summarized in Appendix 4.12. Expression (4.107) can therefore be cast in the simpler form

$$n_c = N_0(T) F_{1/2}(\eta_{fc}),$$

$$N_0(T) = \frac{1}{2\pi^2} \left(\frac{2m_c^* k_b T}{\hbar^2} \right)^{3/2}. \quad (4.109)$$

Here, $\eta_{fc} \equiv (\mu - \varepsilon_c)/(k_b T)$.

In the case of holes similar expressions can be derived, leading to:

$$p_v = P_0(T) F_{1/2}(\eta_{fv}),$$

$$P_0(T) = \frac{1}{2\pi^2} \left(\frac{2m_v^* k_b T}{\hbar^2} \right)^{3/2}, \quad (4.110)$$

with $\eta_{fv} \equiv (\varepsilon_v - \mu)/(k_b T)$. Two limiting cases lead to further simplifications.

Case $\eta_{fc} \gg 1$ or $\eta_{fv} \gg 1$. For electrons and using the properties of Dirac integrals, the condition $\eta_{fc} \rightarrow +\infty$ implies that $n_c \gg N_0(T)$. Considering the results in Appendix 4.12, one finds

$$n_c = \frac{1}{3\pi^2} \left[\frac{2m_c^* (\mu - \varepsilon_c)}{\hbar^2} \right]^{3/2}. \quad (4.111)$$

Comparison with the expression $k_f = (3\pi^2 n)^{1/3}$, leads to the conclusion that this limit recovers the metallic behavior. In a similar manner, the condition $\eta_{fv} \rightarrow +\infty$ for holes leads to:

$$p_v = \frac{1}{3\pi^2} \left[\frac{2m_v^* (\varepsilon_v - \mu)}{\hbar^2} \right]^{3/2}. \quad (4.112)$$

Case $\eta_{fc} < 0$ and $\eta_{fv} < 0$. For electrons and using the properties of Dirac integrals, the condition $\eta_{fc} < 0$ implies $n_c \leq N_0(T)$. In particular, in the $\eta_{fc} \rightarrow -\infty$ limit one finds

$$n_c = N_c(T) e^{\frac{\mu - \varepsilon_c}{k_b T}}$$

$$N_c(T) = 2 \left(\frac{m_c^* k_b T}{2\pi \hbar^2} \right)^{3/2}. \quad (4.113)$$

For holes in the $\eta_{fV} \rightarrow -\infty$ limit, one finds:

$$p_V = P_V(T) e^{\frac{\epsilon_V - \mu}{k_b T}}$$

$$P_V(T) = 2 \left(\frac{m_V^* k_b T}{2\pi \hbar^2} \right)^{\frac{3}{2}}. \quad (4.114)$$

The latter limiting regimes are those which are usually met in semiconductors and correspond, for example, to the situation depicted in Fig. 4.24. Therefore, expressions (4.109) for electrons and (4.110) for holes, along with their simplified versions in the limiting cases $\eta_f \rightarrow \pm\infty$, are one building block for the calculations of thermodynamical and transport quantities in semiconductors. Notice that (4.113) and (4.114) hold for both intrinsic and doped semiconductors. Since the chemical potential μ is an unknown variable, one additional condition is required to determine electron and hole concentrations. Indeed, the chemical potential turns out to be fixed by the doping level. In other words, depending on the impurity concentration within the semiconductor, the chemical potential moves inside the gap, and the electron and hole concentrations change accordingly. The next sections bold this important concept.

To have an idea of typical values for the involved quantities, undoped silicon has $\epsilon_f \simeq 1.1$ eV and $m_c^* \simeq 0.31 m_e$. The concentration of thermally excited electrons at room temperature $T \simeq 300$ K is easily calculated from (4.113). After considering that in the intrinsic case $\mu = \epsilon_V + E_g/2$ as detailed in the next section, one has:

$$n_c = 2 \left(\frac{0.31 \cdot 9.11 \cdot 10^{-28} \cdot 1.38 \cdot 10^{-16} \cdot 300}{2\pi \cdot (1.054 \cdot 10^{-27})^2} \right)^{\frac{3}{2}}$$

$$\cdot e^{\frac{-0.55 \cdot 1.6 \cdot 10^{-12}}{1.38 \cdot 10^{-16} \cdot 300}} \simeq 4.34 \cdot 10^{18} \text{ cm}^{-3} \cdot e^{-21.26}$$

$$\simeq 2.54 \cdot 10^9 \text{ cm}^{-3}$$

Typical electron densities in metals can be as large as 10^{22} cm^{-3} .

Quick Questions

Q26. Diamond, which is an insulator with $E_g \simeq 6$ eV would be useless as

a transport channel in microelectronic devices, contrary to Si. Why?

Answer - Reverting to (4.113), it easily seen that at room temperature the ratio of silicon-to-carbon thermally excited electrons in the conduction band would be of the order of

$$\frac{n_c(\text{Si})}{n_c(\text{C})} \simeq \frac{e^{-1.1 \text{ eV}/2k_b T}}{e^{-6.0 \text{ eV}/2k_b T}} \simeq e^{2.45 \text{ eV}/0.025 \text{ eV}} = e^{98} \simeq 10^{42},$$

where the ratio of the effective masses of silicon and diamond has been neglected. That is to say, at room temperature a carrier concentration 10^{42} times smaller than that in silicon would be measured in diamond, leading in fact to no detectable or quite negligible electric current.

4.4.2 Intrinsic semiconductors

Definition

A semiconductor is defined as intrinsic if it does not contain impurities.

In this case, one has $n_c = p_v = n_i$ and therefore

$$N_0(T)F_{1/2}\left(\frac{\mu_i - \varepsilon_c}{k_b T}\right) = P_0(T)F_{1/2}\left(\frac{\varepsilon_v - \mu_i}{k_b T}\right). \quad (4.115)$$

Equation (4.115) can be used to determine the chemical potential μ_i in the intrinsic case. Using the asymptotic relations $F_{1/2}(x \rightarrow -\infty)$ (4.113) and (4.114), one finds that:

$$N_c(T)e^{\frac{\mu_i - \varepsilon_c}{k_b T}} = P_v(T)e^{\frac{\varepsilon_v - \mu_i}{k_b T}},$$

where the chemical potential can be extracted from, and precisely:

$$\mu_i = \frac{\varepsilon_c + \varepsilon_v}{2} - \frac{3k_b T}{4} \ln\left(\frac{m_c^*}{m_v^*}\right). \quad (4.116)$$

In particular, $\mu_i(T = 0 \text{ K}) = (\varepsilon_c + \varepsilon_v)/2$: that is, at zero temperature the chemical potential lies in the middle of the gap. Besides, in most semiconductor materials the condition $m_v^* > m_c^*$ occurs. For these materials μ is an increasing function of the temperature. The number of electrons in the conduction band and of holes in the valence band is given by the expression:

$$n_i = \sqrt{P_v(T)N_c(T)}e^{-\frac{E_g}{2k_b T}} = 2(m_c^*m_v^*)^{3/4} \left(\frac{k_b T}{2\pi\hbar^2}\right)^{3/2} e^{-\frac{E_g}{2k_b T}}, \quad (4.117)$$

with $E_g = \varepsilon_c - \varepsilon_v$.

The procedure to calculate the chemical potential can be coded into these three simple steps:

Procedure

Step 1. Write the expression for the total number of particles, using the appropriate quantum distribution function in terms of the energy levels of the system, its chemical potential, and $k_b T$.

Step 2. Identify the expression for the energy levels to be inserted in Step 1.

Step 3. Invert the relation between the total number of particles and the chemical potential, to determine the latter. This step might require approximations to carry out analytical calculations, or even some numerical work.

The procedure is now being applied to the following example, where the valence and conduction bands are mimicked by two single levels ε_1 and ε_2 , respectively, with degeneracy. Further examples are given, for impurity states, in Sec. 4.4.4.

Examples

Consider a system composed of only two levels with energies ε_1 and ε_2 , each of them with degeneracy d_1 and d_2 , respectively, and a number of electrons $N = d_1$. The chemical potential μ can be calculated as follows. Since the total number of particles is

$$N = \frac{d_1}{e^{\beta(\varepsilon_1 - \mu)} + 1} + \frac{d_2}{e^{\beta(\varepsilon_2 - \mu)} + 1},$$

it follows that

$$1 = \frac{1}{x + 1} + \frac{d}{e^{\beta\Delta\varepsilon}x + 1}$$

with $\beta = k_b T$, $x = e^{\beta(\varepsilon_1 - \mu)}$, $d = d_2/d_1$, and $\Delta\varepsilon = \varepsilon_2 - \varepsilon_1$. After some algebra, the equation above can be turned into a quadratic algebraic equation for the unknown x . The physically meaningful solution is

$$e^{\beta(\varepsilon_1 - \mu)} = \frac{d - 1 + \sqrt{(d - 1)^2 + 4de^{\beta\Delta\varepsilon}}}{2e^{\beta\Delta\varepsilon}}.$$

From this expression, μ can be extracted. In the limiting case with $e^{\beta\Delta\varepsilon} \gg (d - 1)^2/(4d)$, one has

$$\mu = \frac{\varepsilon_1 + \varepsilon_2}{2} - \frac{k_b T}{2} \ln d.$$

In the opposite limit with $e^{\beta\Delta\varepsilon} \ll (d - 1)^2/(4d)$, one finds instead:

$$\mu = \varepsilon_2 - k_b T \ln(d - 1) \text{ for } d > 1 \quad (4.118)$$

$$\mu = \varepsilon_1 - k_b T \ln \frac{d}{1 - d} \text{ for } d < 1. \quad (4.119)$$

Finally, if $d = 1$ then $\mu = (\varepsilon_1 + \varepsilon_2)/2$.

4.4.3 Extrinsic semiconductors

Definition

A semiconductor is defined as extrinsic if it does contain impurities able to provide the material with extra electrons or holes.

Thus, an excess or a vacancy of electrons or holes takes place with respect to the intrinsic case, so that $n_c \neq n_i$ and $p_v \neq n_i$.

In the case of doped semiconductors, expressions (4.109) and (4.110) are still valid. Therefore,

$$\begin{aligned} n_c p_v &= N_0(T) F_{1/2}(\eta_{fc}) P_0(T) F_{1/2}(\eta_{fv}) \\ &= n_i^2 \frac{F_{1/2}(\eta_{fc}) F_{1/2}(\eta_{fv})}{F_{1/2}(\eta_{fc}^{(i)}) F_{1/2}(\eta_{fv}^{(i)})}. \end{aligned} \quad (4.120)$$

Here, $\eta_{fc} = \beta(\mu - \varepsilon_c)$, $\eta_{fv} = \beta(\varepsilon_v - \mu)$, $\eta_{fc}^{(i)} = \beta(\mu_i - \varepsilon_c)$ and $\eta_{fv}^{(i)} = \beta(\varepsilon_v - \mu_i)$, in terms of the chemical potential μ_i in the intrinsic case. Expression (4.120) implies that $n_c p_v \leq n_i^2$, as it can be numerically checked. Indeed, when the asymptotic relations for the Dirac integrals can be used along with $\mu > \varepsilon_c$, one finds that:

$$n_c p_v = n_i^2 \frac{4}{3\sqrt{\pi}} e^{-\frac{\mu - \varepsilon_v - E_g}{k_b T}} \left(\frac{\mu - \varepsilon_c}{k_b T} \right)^{\frac{3}{2}}. \quad (4.121)$$

The factor in front of n_i^2 is smaller than unity. A similar reasoning holds in the case with $\varepsilon_v > \mu$. Finally, when $\varepsilon_v < \mu < \varepsilon_c$, the use of asymptotic relations yields:

$$n_c p_v = n_i^2. \quad (4.122)$$

In essence,

Properties

P1. In extrinsic semiconductors $n_c p_v$ cannot ever exceed n_i^2 whatever the position and the number of doping-induced impurities states might be. This is a formulation of the mass-action law.

The parameter measuring the degree of extrinsic behavior is conveniently introduced as $\Delta n \equiv n_c - p_v$. In the case in which (4.122) holds, one has

$$\begin{aligned}\frac{p_V}{n_i} &= \frac{1}{2} \left(\sqrt{\left(\frac{\Delta n}{n_i}\right)^2 + 4} - \frac{\Delta n}{n_i} \right) \\ \frac{n_c}{n_i} &= \frac{1}{2} \left(\sqrt{\left(\frac{\Delta n}{n_i}\right)^2 + 4} + \frac{\Delta n}{n_i} \right).\end{aligned}\quad (4.123)$$

Two opposite limiting behaviors are significant here. If $|\Delta n|/n_i \ll 1$ the intrinsic regime is recovered. If instead $|\Delta n|/n_i \gg 1$, one obtains:

$$\begin{aligned}\frac{n_c}{n_i} &\sim \frac{\Delta n}{n_i}, \quad \frac{p_V}{n_i} \sim \frac{n_i}{\Delta n}, \quad \text{for } \Delta n > 0 \\ \frac{p_V}{n_i} &\sim \frac{|\Delta n|}{n_i}, \quad \frac{n_c}{n_i} \sim \frac{n_i}{|\Delta n|}, \quad \text{for } \Delta n < 0.\end{aligned}\quad (4.124)$$

Since n_i is exponentially decreasing with temperature, at given fixed doping a semiconductor may occur to be in an intrinsic regime at high temperatures and extrinsic regime at lower ones.

Definition

Majority (minority) carriers are defined as the outnumbering carrier type, that is electrons if $n_c > p_V$ ($n_c < p_V$) and holes viceversa.

Numerical analysis show that at low temperatures the number of majority carriers varies slightly with temperature, at variance with minority carriers.

4.4.4 Donor and acceptor states

This section illustrates how extrinsic semiconductors can be realized in practice. Consider for example a Si crystal: substitution of one Si atom with a species from the fifth column in the periodic table, e.g. As, determines the presence of an extra valence electron. This electron turns out to be only weakly bonded with the binding energy $\simeq 10^{-2}$ eV, in that the host atom can be easily ionized, as e.g. due to thermal excitations. After ionization, the extra electron freely moves throughout the crystal and can contribute to the conduction processes, whereas the As impurity atom remains positively charged. Thus, in the vicinity of the impurity a Coulomb potential is generated. Due to the tendency of donating one electron to the host lattice, impurities like phosphorous are usually called donors. If the substitution were performed with an element from the third column, e.g. B, the impurity atom now has one missing electron to complete the four valence bonds. It turns out that a modest energy is required for a valence band electron to complete the unsatisfied valence. In other words, thermal excitation of an electron from the valence band can take place, resulting in a negatively ionized impurity and a hole left behind in the valence band.

Such hole can contribute to conduction processes as well. Again, in the vicinity of the impurity a Coulomb potential is set up. In this case, due to the tendency to trap an electron from the host lattice, impurities like boron are called acceptors. This ionization processes are schematically reported in the top panels of [Fig. 4.25](#).

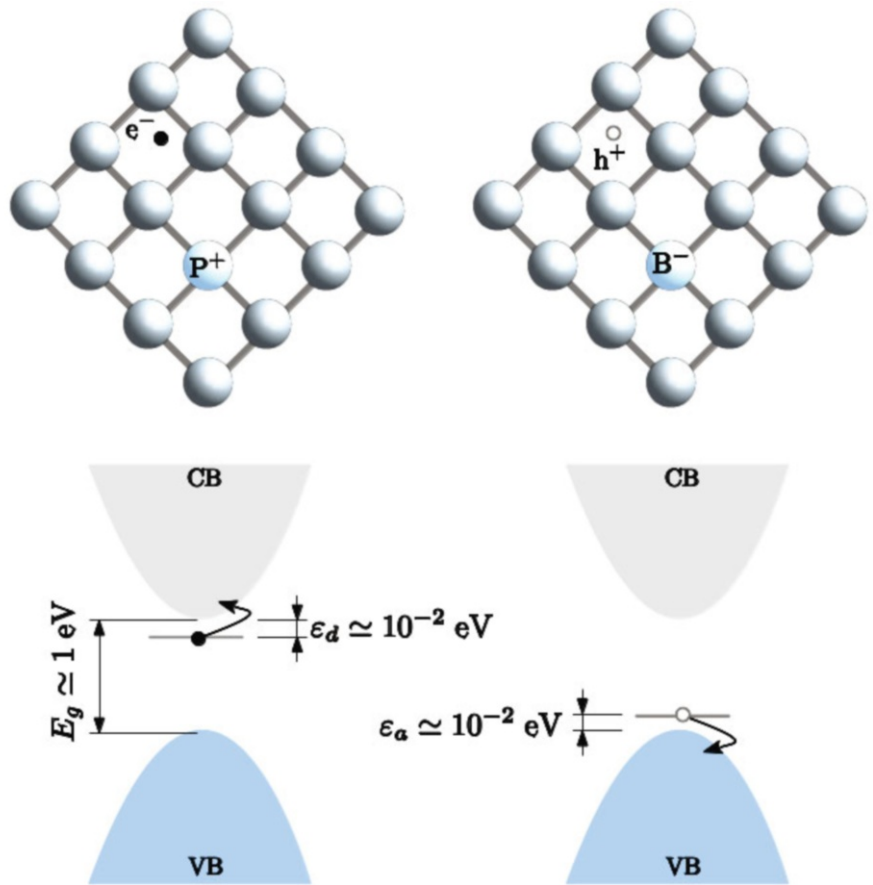


Fig. 4.25 Transport behavior in semiconductors. Top panels: schematic representation of a IV group semiconductor, such as e.g. Si or Ge, in the presence of an impurity. Left panels: donor impurities, such as e.g. P or As. Right panels: acceptor impurities, such as e.g. B or Al. Bottom panels: the modification of semiconductor electronic levels after the introduction of a donor (left part) or acceptor (right part) into the host lattice. The impurity level is occupied [or empty] at $T = 0$ K in the case of a donor [or acceptor]. The arrows indicate the thermal excitation of the donor electron [or acceptor hole]. See text for details

One remark is useful here: the extra electron [hole], provided to the conduction [valence] band by thermal ionization of the impurity, is in excess with respect to those which may have been excited by thermal excitation of electrons from the valence to the conduction band across the energy gap. Thus, a semiconductor doped

with donor impurities has more negative charge carriers than the intrinsic one and is called an *n*-type semiconductor. On the other hand, a semiconductor doped with acceptor impurities has more positive charge carriers than the intrinsic one and is called a *p*-type semiconductor.

A more quantitative discussion of the donor and acceptor states can be performed after using a few concepts parallel to those used in Chap. 2 to describe the excitonic effect. There, extra electrons are promoted from the valence to the conduction band. Here, one considers the case in which an extra electron [hole] comes about after introducing a substitutional donor [acceptor] impurity in the crystal, having a different valence from that of the atoms composing the hosting crystal.

In principle, understanding of this system should be better faced by studying the problem of $N + 1$ or $N - 1$ electrons in the combined periodic and impurity potentials. On the other hand, one can show that under general enough conditions this problem can be considered equivalent to that of one single extra electron in an either attractive or repulsive potential, as shown in the top panels of Fig. 4.25. The latter has Coulomb-like, long-range behavior and at short distances depends on the nature of the impurity and of the hosting crystal. In addition, it shifts downwards the energy of a donor state with respect to the conduction-band bottom, and it shifts upwards the energy of an acceptor state with respect to the valence-band top.

For example, a donor state can be described by the hydrogen-like equation

$$\left(\epsilon_c - \frac{\hbar^2}{2m_c^*} \nabla^2 - \frac{e^2}{\epsilon_0 r} \right) \psi(\mathbf{r}) = E_d \psi(\mathbf{r}), \quad (4.125)$$

with r the distance between electron and impurity and ϵ_0 the static dielectric constant. In this equation, the ionized donor impurity acts as a positive point charge which attracts the extra-electron, as in the top panel of Fig. 4.25. Equation (4.125) is valid if the wavefunction $\psi(\mathbf{r})$ extends over a size larger than a lattice distance. As a consequence, when expanding $\psi(\mathbf{r})$ in Bloch functions, only those in the vicinity of the bottom of the conduction band bring a significant contribution. One then finds for the donor energy E_d and the effective Bohr radius a_b^{eff} the expressions

$$\begin{aligned} E_d &= \epsilon_c + \epsilon_{d,n} = \epsilon_c - \frac{m_c^* e^4}{2\hbar^2 \epsilon_0^2 n^2} \\ a_b^{eff} &= \frac{\epsilon_0 \hbar^2}{m_c^* e^2}. \end{aligned} \quad (4.126)$$

It is evident that electronic states of donor impurities are bound, their energy being determined by the effective mass related to the conduction band and by the static dielectric constant of the material. Since in a semiconductor typically $\epsilon_0 \sim 10$ and $m_c^* \sim 0.1m_e$, binding energies are of the order of ~ 10 meV, that is about $1/1000$ of that of hydrogen atom in vacuum. Correspondingly, typical values of the Bohr radii are $a_b^{eff} \sim 50 \text{ \AA}$, that is about 100 times larger than the Bohr radius of the hydrogen atom in vacuum. These numbers justify a posteriori the starting assumptions.

In the case of acceptor impurities instead, (4.125) is to be replaced by

$$\left(\varepsilon_V + \frac{\hbar^2}{2 |m_V^*|} \nabla^2 + \frac{e^2}{\varepsilon_0 r} \right) \psi(\mathbf{r}) = E_a \psi(\mathbf{r}), \quad (4.127)$$

with m_V^* the valence-band effective mass as calculated in the vicinity of the valence-band top. Notice that at variance with the case of donor impurities, the effective mass here is negative and the potential repulsive. With the same assumptions than in the donor case, one finds

$$E_a = \varepsilon_V + \varepsilon_{a,n} = \varepsilon_V + \frac{|m_V^*| e^4}{2 \hbar^2 \varepsilon_0^2 n^2}$$

$$a_b^{eff} = \frac{\varepsilon_0 \hbar^2}{|m_V^*| e^2}. \quad (4.128)$$

The energies of the impurities states (4.126) and (4.128) are calculated within an approximate model. Deviations may occur due in turn to deviations of the impurity potential from the Coulomb-like form, leading to a shift in the chemical potential, and to the presence of multiple minima or maxima in the conduction or valence bands, respectively, at almost degenerate or degenerate energies. Additional deviations may be originated by particularly strong values of the impurity potential, such as those that could characterize interstitial impurities. The impurity state is thus deeply bound and its energy depends on the characteristics of the whole, either valence or conduction, band. The bottom panels of Fig. 4.25 display the position of the donor and acceptor levels, along with the related binding energies ε_d in (4.126) and ε_a in (4.128).

Once the energies and wavefunctions of the donor and acceptor states are known, for statistical calculations an expression is needed for the probability that an electron occupy a donor state or a hole occupy an acceptor state. The detailed derivation is reported in Appendix 4.13, where it is assumed that no two electrons or holes can occupy the same impurity state. The result can be expressed as

$$\langle n \rangle = P_e(E_d) = \frac{1}{1 + \frac{1}{2} e^{\beta(E_d - \mu)}} \quad (4.129)$$

and

$$\langle p \rangle = P_h(E_a) = \frac{1}{1 + \frac{1}{2} e^{\beta(\mu - E_a)}}. \quad (4.130)$$

The factor 1/2 guarantees that the state be singly occupied, according to the above assumption.

Expressions (4.129) and (4.130) still contain the unknown chemical potential μ . This is fixed by imposing the condition the total number of charged particles has to be zero to ensure system neutrality. One has:

$$n_c - p_v = N_r = \sum_j N_{d_j} (1 - P_e(E_{d_j})) - \sum_k N_{a_k} P_e(E_{a_k}). \quad (4.131)$$

Here, N_{d_j} and N_{a_k} represent the number of donors of species j and of acceptors of species k , respectively. In addition, $N_{d_j}(1 - P_e(E_{d_j}))$ represents the number of ionized donors while $N_{a_k}P_e(E_{a_k})$ the number of acceptors with one extra electron. Equations (4.129) and (4.130) evidence that donors with energies $\epsilon_c + \epsilon_{d_j} \ll \mu$ and acceptors with energies $\epsilon_v + \epsilon_{a_k} \gg \mu$ yield negligible contribution to the right-hand side of (4.131). Their contribution is instead relevant whenever the conditions $\epsilon_c + \epsilon_{d_j} \geq \mu$ and $\epsilon_v + \epsilon_{a_k} \leq \mu$ are met, respectively. In conclusion, the right-hand side of (4.131) is significantly nonzero when μ gets closer to the energy of a donor or acceptor state.

The solution of (4.131) can be easily discussed under the two following simplifying assumptions: first, that only one donor or acceptor species be present with $N_d > N_a$; and, second, that the gap be so large that $p_v \sim 0$ and $\epsilon_v + \epsilon_a \ll \mu$, and thus $P_e(E_a) \sim 1$. The equation thus becomes:

$$n_c = N_d(1 - P_e(E_d)) - N_a. \quad (4.132)$$

In any event, the solution of (4.131) or of its simplified version (4.132) yields the chemical potential μ as a function of temperature, once the other governing parameters have been either fixed or calculated.

In particular, the following procedure can be conveniently used to determine the donor or acceptor problem [the square bracketed quantities refer to the acceptor problem]:

Procedure

Step 1. Get ϵ_0 and $m_c^* [m_v^*]$ for the selected material, and fix the number of donors N_d [acceptors N_a].

Step 2. Solve (4.126) [(4.128)] to determine ϵ_d [ϵ_a].

Step 3. Use (4.131) or its simplified version (4.132) to calculate the chemical potential $\mu(T)$ as a function of temperature.

Step 4. Use (4.109) [(4.110)] or alternative simplified versions to calculate n_c [p_v] and the number of ionized impurities if required.

Instead of entering detailed analytic calculations about selected cases, the qualitative behavior of $\mu(T)$ is better sketched in Fig. 4.26. A case with $N_d > N_a$ is illustrated in panel (a) and a case with $N_a > N_d$ in panel (b). The n-doped case in panel (a) is now discussed in more detail, similar reasoning holding for the p-doped case depicted in panel (b). Fig. 4.26 (a) makes evident that at very low temperatures all acceptor states are occupied by the electrons along with part of the donor states, the limiting value of μ being dictated by the position of ϵ_d and ϵ_c . While the temperature increases, donor states are seen to ionize more easily than acceptor ones, this tendency being progressively neater as long as the difference $N_d - N_a$ gets tinier:

in fact, the chemical potential is lowering while it approaches the gap center. Further temperature increases induce the electrons of acceptor states to move up to the conduction band: the chemical potential lowers further towards the energy position of acceptor states, until the latter are almost completely ionized. The final position of μ is determined by the actual values of ϵ_V and ϵ_a .

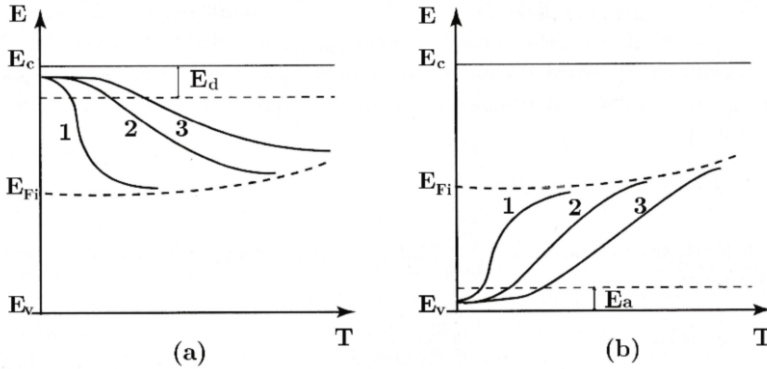


Fig. 4.26 Transport behavior of extrinsic semiconductors. Qualitative behavior of the chemical potential as a function of temperature. (a) n-doped semiconductor with $N_d > N_a$, (b) p-doped semiconductor with $N_a > N_d$. Increasing values of the labels from 1 to 3 refer to increasing values of impurity densities. The donor E_d and acceptor E_a energies are indicated

The following examples give a scheme of as much real possible situations, in which the valence and conduction bands are simplified as two single levels ϵ_1 and ϵ_2 , respectively, with given degeneracy and the impurity state as an additional new level with lower degeneracy, inserted between the first two. The chemical potential is calculated as as function of the temperature and degeneracy of the levels.

Examples

Consider the two-level system of the previous example. Introduce a third level ϵ_3 in the interval (ϵ_1, ϵ_2) , thus lying in between the first two. In addition, the following relation $\epsilon_d = \epsilon_2 - \epsilon_3 \ll \epsilon_2 - \epsilon_1$ holds: that is, the third level is much closer to ϵ_2 than to ϵ_1 . Assume that the third level be degenerate with degeneracy $d_3 \ll d_1$, and set $d' = d_3/d_1$, that $d_2 = d_1$, with the total number of electrons being $N = d_1 + d_3$. The chemical potential is now being determined in the limits $(\epsilon_2 - \epsilon_1) \gg k_b T$ and $\epsilon_d \gg k_b T$. At zero temperature all the energy states ϵ_1 and ϵ_3 are occupied: under the conditions $(\epsilon_2 - \epsilon_1) \gg k_b T$ and $\epsilon_d \gg k_b T$, it is then reasonable to suppose that $(\mu - \epsilon_1)/(k_b T) \gg 1$. Considering

$$N = \frac{d_1}{e^{\beta(\epsilon_1 - \mu)} + 1} + \frac{d_1}{e^{\beta(\epsilon_2 - \mu)} + 1} + \frac{d_3}{e^{\beta(\epsilon_3 - \mu)} + 1}, \quad (4.133)$$

the first term on the right-hand side results to be equal to d_1 and the relation above can be cast into

$$1 + d' = 1 + \frac{1}{e^{\beta \varepsilon_d} x + 1} + \frac{d'}{x + 1},$$

with $x \equiv e^{[\beta(\varepsilon_3 - \mu)]}$. The resulting quadratic algebraic equation can be solved for the unknown x , yielding

$$x = \frac{1 - d' + \sqrt{(1 - d')^2 + 4d'e^{\beta \varepsilon_d}}}{2d'e^{\beta \varepsilon_d}}.$$

Since d' is in most realistic cases a very small quantity, the values of x are very sensitive to changes of $r' = d'e^{\beta \varepsilon_d}$. In the limit $2r' \gg 1$ one then finds:

$$\mu = \varepsilon_2 - \frac{\varepsilon_d}{2} + \frac{k_b T}{2} \ln d'.$$

In the opposite limit $2r' \ll 1$ instead, one has:

$$\mu = \varepsilon_2 - k_b T \ln \frac{1 - d'}{d'}.$$

Examples

Consider once again the same system of the previous example, with the difference now that the assumption $\varepsilon_d \gg k_b T$ be substituted by the opposite one $\varepsilon_d \ll k_b T$. The temperature is high enough for the electrons in state ε_3 to be considered all excited. Under these new conditions, the equation to be solved becomes

$$1 + d' = \frac{1}{x + 1} + \frac{1}{e^{\beta \Delta \varepsilon} x + 1} + \frac{d'}{e^{-\beta \varepsilon_d} e^{\beta \Delta \varepsilon} x + 1},$$

with $x = e^{[\beta(\varepsilon_1 - \mu)]}$ and $\Delta \varepsilon = \varepsilon_2 - \varepsilon_1$. Since $\varepsilon_d \ll k_b T$, the equation above can be cast in the approximated form

$$1 + d' = \frac{1}{x + 1} + \frac{1 + d'}{e^{\beta \Delta \varepsilon} x + 1},$$

whose solution is

$$x = \frac{-d'e^{\beta \Delta \varepsilon} + \sqrt{d'^2 e^{2\beta \Delta \varepsilon} + 4(1 + d')e^{\beta \Delta \varepsilon}}}{2(1 + d')e^{\beta \Delta \varepsilon}}.$$

The quantity $r = d'e^{\frac{\beta \Delta \varepsilon}{2}}$ again plays a crucial role. In particular, if $r \gg 2$ one finds that:

$$\mu = \varepsilon_2 + k_b T \ln d'.$$

If $r \ll 2$ instead, one finds

$$\mu = \varepsilon_1 + \frac{\Delta\varepsilon}{2} + \frac{k_b T}{2} \ln(1 + d'),$$

which has the same form as that obtained for intrinsic semiconductors.

Examples

Consider now the case of an electronic state with energy ε_3 close to the lowest energy state. This is characterized by the following conditions: $\varepsilon_a = \varepsilon_3 - \varepsilon_1 \ll \varepsilon_2 - \varepsilon_1$, $\varepsilon_3 = \varepsilon_1 + \varepsilon_a$, $N = d_1$ and $\varepsilon_a \gg k_b T$. The electronic state ε_3 is essentially empty, therefore the chemical potential μ lies below ε_3 . Thus, the energy state with ε_2 is certainly empty and (4.133) becomes:

$$1 = \frac{1}{x+1} + \frac{d'}{e^{\beta\varepsilon_a} x + 1},$$

whose solution is

$$x = \frac{-(1-d') + \sqrt{(1-d')^2 + 4d'e^{\beta\varepsilon_a}}}{2e^{\beta\varepsilon_a}}.$$

If $\sqrt{4d'e^{\beta\varepsilon_a}} \gg 1$, one has

$$\mu = \varepsilon_1 + \frac{\varepsilon_a}{2} - \frac{k_b T}{2} \ln d',$$

whereas if $\sqrt{4d'e^{\beta\varepsilon_a}} \ll 1$ one finds:

$$\mu = \varepsilon_1 + \varepsilon_a - k_b T \ln \frac{d'}{1-d'}.$$

Examples

Consider finally the latter example, now with the condition $\varepsilon_a \ll k_b T$. In a manner similar to the second example, one finds

$$1 = \frac{1+d'}{x+1} + \frac{1}{e^{\beta\Delta\varepsilon} x + 1},$$

whose solution is

$$x = \frac{d'e^{\beta\Delta\varepsilon} + \sqrt{d'^2 e^{2\beta\Delta\varepsilon} + 4(1+d')e^{\beta\Delta\varepsilon}}}{2e^{\beta\Delta\varepsilon}}.$$

If $d'e^{\frac{\beta\Delta\varepsilon}{2}} \gg 2$, one has:

$$\mu = \varepsilon_1 - k_b T \ln d',$$

while if $d' e^{\frac{\beta \Delta \varepsilon}{2}} \ll 2$ one obtains:

$$\mu = \varepsilon_1 + \frac{\Delta \varepsilon}{2} - \frac{k_b T}{2} \ln(1 + d').$$

Fig. 4.27 illustrates the qualitative behavior of $n_c(1/T)$ for n-doped Si as it is extracted from experimental data [21]. Different regimes are distinguished. In the $T \rightarrow 0$ limit, $n_c \rightarrow 0$ due to the fact that electrons tend to leave the conduction band occupying back states of both valence band and impurities: this is referred to as freezing region. The intermediate flat region at higher temperatures represents the situation in which all the electrons have moved from the impurity to the conduction-band states: when this occurs, a sort of saturation takes place in which the mechanism for which electrons are promoted from the valence to the conduction band is not immediately switched on, the gap energy being $E_g = \varepsilon_c - \varepsilon_v \gg |\varepsilon_d|$. In the higher temperature region, the latter mechanism switches on and the intrinsic regime is reached, where it has been assumed that $N_d \ll N_A$, the Avogadro number.

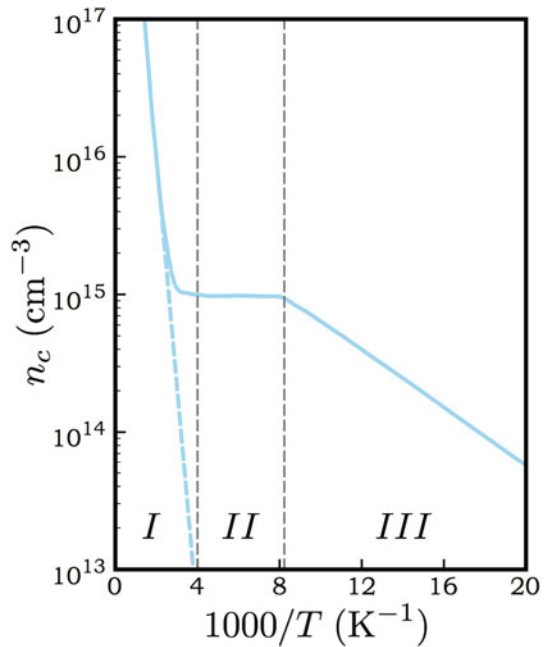


Fig. 4.27 Transport behavior of extrinsic semiconductors. Behavior of n_c as a function of $1000/T$ in K^{-1} as extracted from experimental data for n-doped Si with $N_d = 10^{15} \text{ cm}^{-3}$. Three different regions are visible where different generation mechanisms prevail, as discussed in the main text [21]

To conclude, the following concept is especially relevant:

Concept

Doping in semiconductors is a quite versatile tool to tune the thermodynamic and transport properties of the material by tuning the position of the chemical potential and the number of carriers (electrons and holes). This concept can be extraordinarily useful for applications.

The explicit calculation of μ in doped materials has not been performed: Fig. 4.26 shows the qualitative behavior of such quantity as a function of temperature.

4.5 Superconductors

In Chap. 2 the ground state and electronic excitations of superconductors have been derived first following a phenomenological approach and then within the Bardeen, Cooper, and Schrieffer theoretical framework. The crucial finding is that the superconducting ground state is characterized by an electronic structure $E_{\mathbf{k}} = \sqrt{(\epsilon - \mu)^2 + \Delta_0^2}$ with a finite gap Δ_0 for the elementary excitations. The gap is a consequence of pairing between particles originated by a whatever weak attractive interaction, that correlates electrons with opposite momenta and spins lying within a slice tailored around the Fermi surface as thick as the typical energy provided by the excitations that mediate the pairing. In the new ground state, the pairs are entangled among themselves as their size is larger than interparticle spacing. In ordinary metallic superconductors, these are acoustical phonons driving an effective attraction between electrons via a weak electron-phonon interaction, as described in Chap. 3. Under these conditions, the BCS theory predicts an exponential behavior of the gap at zero temperature, $\Delta_0 \simeq 2\hbar\omega_D e^{-2/[g(\epsilon_f)V]}$ in terms of the typical phonon energy $\hbar\omega_D \ll \epsilon_f$, of the electron-phonon interaction strength V and the density of electronic states $g(\epsilon_f)$ at the Fermi level. These results have deep consequences in the transport behavior of superconductors, as discussed below.

4.5.1 Conductivity and superconducting flow

Resistivity. First, the most striking consequence of the existence of the gap below T_c , is that no states are available within an energy interval $2\Delta_0$ above the superconducting ground state, into which electrons can scatter into via either one of the possible scattering mechanisms leading to resistivity. Neither phonon quanta of sufficient energy at such low temperatures, nor electron-electron scattering processes of large enough energy are available to overcome the gap:

Concept

The correlated ground state behaves in coherent manner, the system behaves as if it were one single particle not having though an available state where to scatter into. This explains why below T_c the resistivity suddenly drops to vanishing values.

Josephson effect. The existence of a global phase characterizing the superfluid coherence and determining the superfluid velocity bears one more peculiar consequence. Consider a superconductor in the phenomenological model, characterized by a macroscopic occupation of the condensate wavefunction, so that $\psi(\mathbf{r}, t) = \sqrt{n_s(\mathbf{r}, t)} e^{i\phi(\mathbf{r}, t)}$. The wavefunction satisfies the Schrödinger equation

$$i\hbar \partial \psi(\mathbf{r}, t) / \partial t = -(\hbar^2 / 2m_s) \nabla^2 \psi(\mathbf{r}, t) + \mu \psi(\mathbf{r}, t).$$

Assuming that $\psi(\mathbf{r}, t)$ is a slowly varying function of space, the kinetic term can be neglected and one finds that $\hbar \partial \phi(\mathbf{r}, t) / \partial t \simeq -\mu$, that is the phase of the wave function evolves linearly in time, $\phi \simeq -\mu t / \hbar$, and $\psi(\mathbf{r}, t) \approx e^{-i\mu t / \hbar}$.

Consider now the case in which two such superconductors form a junction and represented by a left $\psi_L(\mathbf{r}, t)$ and a right $\psi_R(\mathbf{r}, t)$ wavefunction. From now on, the subscripts L and R refer to variables belonging to the superconductor at the left and right sides of the junction, respectively. If the junction barrier is sufficiently thin to allow tunneling of charge carriers, one can write on phenomenological grounds the following coupled equations for the two joined superconductors:

$$\begin{aligned} i\hbar \frac{\partial \psi_L(t)}{\partial t} &= \mu_L \psi_L(t) + C \psi_R(t) \\ i\hbar \frac{\partial \psi_R(t)}{\partial t} &= \mu_R \psi_R(t) + C \psi_L(t), \end{aligned} \quad (4.134)$$

in the limit of a weak, linear coupling between the two junctions. The particles current in the junction is $J = \partial n_R / \partial t = -\partial n_L / \partial t$. Separation of the real and imaginary parts of the equations yields:

$$\begin{aligned} -\frac{\partial n_{sL}(t)}{\partial t} &= \frac{\partial n_{sR}(t)}{\partial t} = \frac{2C}{\hbar} \sqrt{n_{sL} n_{sR}} \sin[\phi(t)] \\ \frac{\partial \phi(t)}{\partial t} &= -\frac{\mu_L - \mu_R}{\hbar}, \end{aligned} \quad (4.135)$$

in terms of the phase difference $\phi = \phi_L - \phi_R$ and assuming that $n_R \simeq n_L$. The two equations above state together that a current $I(t)$ flows across the junction, that is $I(t) = I_0 \sin[\phi(t)]$ with $I_0 = 2e_s C \sqrt{n_{sL} n_{sR}} / \hbar$ and $\phi(t) = \phi_0 - (\mu_L - \mu_R)t / \hbar$ after integration, driven by the difference of chemical potentials $(\mu_L - \mu_R)$ across the junction. In essence

Concept

Even in the absence of applied voltage difference, a tunneling supercurrent of pairs $I = I_0 \sin[\phi]$ sets in across the junction due to the difference in the two global superconductor phases. This is called Josephson current.

If at the junction a current $I < I_0$ flows, the phases adjust themselves so to generate a supercurrent flow, that is a current of superconducting pairs, in the absence of a voltage drop. If $I > I_0$ instead, a voltage drop eV is generated and a current flows only if $eV \geq 2\Delta_0$, that is therefore carried by normal electrons.

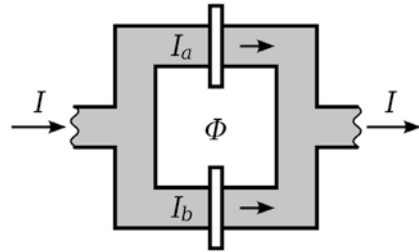


Fig. 4.28 Transport behavior in superconductors. Scheme of a SQUID, made of two superconductor slabs S1 and S2 faced to form two junctions a and b . See text for details

The Josephson effect can be exploited to perform precision measurements of external forces and fundamental constants, especially when combined with flux quantization. Consider indeed two superconductor slabs S1 and S2 that are faced in a way to establish two junctions a and b between them at two different points, as in Fig. 4.28. A closed circuit can be therefore devised, one half of which lying within S1 and the second half within S2, the two halves joined at a and b . Such a circuit is characterized by the total phase difference

$$[\phi_{S1}(a) - \phi_{S1}(b)] - [\phi_{S2}(a) - \phi_{S2}(b)] = e_s / (\hbar c) \left[\int_{S1,a}^{S1,b} \mathbf{A} \cdot d\ell - \int_{S2,a}^{S2,b} \mathbf{A} \cdot d\ell \right] = e_s \Phi / (\hbar c),$$

where the last equality results from definition of flux Φ through a closed circuit, as seen in Sec. 2.7. The current is mainly given by the supercurrent at the two junctions, so that $I \simeq I_0 \{ \sin[\phi_{S1}(a) - \phi_{S2}(a)] + \sin[\phi_{S1}(b) - \phi_{S2}(b)] \}$ or

$$I \simeq I_0 \{ \sin[\phi_{S1}(a) - \phi_{S2}(a) + \phi_{S1}(b) - \phi_{S2}(b)] / 2 \} \cos \left(\frac{e_s \Phi}{2\hbar c} \right).$$

In essence

Concept

Maximum currents can be obtained after appropriate tuning of the magnetic flux trapped by the device or else the external magnetic field.

Such a model schematizes the working concept of a so-called Superconducting Quantum Interference Device or SQUID. SQUIDS have fundamental applications in measuring tiny magnetic fields, such as those involved in medical imaging of the brain, and in the generation of electromagnetic fields at a wanted, well-defined, frequency with important applications in microelectronics and telecommunications.

4.5.2 Specific heat

The existence of gap in the electronic structure explains the behavior of the specific heat in normal superconductors. As shown in Fig. 2.46 (c), experiments are consistent with a specific heat that exponentially increases while $T \rightarrow T_c$ and then drops for $T > T_c$ to the linear behavior typical of metals. In fact, if a gap is present in the excitation spectrum, particles at the Fermi level are expected to be thermally activated according to the expression $e^{-\beta\Delta(T)}$: while $T \rightarrow T_c$, $\Delta(T) \rightarrow 0$ and the activation exponentially increases. Since the specific heat is proportional to the number of excitations, one has $c_V \propto e^{-\beta\Delta(T)}$.

4.5.3 Magnetic behavior

The ground state of conventional superconductors is characterized by a singlet pairing, where the Cooper pairs are coupled with opposite momenta and spins. On average then, the behavior of a superconductor is similar to that of a perfect diamagnet, that is able to completely expel the magnetic field from its bulk. While temperature increases from zero to T_c , pairs are progressively broken up while the superconducting gap $\Delta(T)$ drops to zero as $\Delta_0\sqrt{1 - (T/T_c)}$ close to T_c , so that perfect diamagnetism is progressively lost. The relationship between critical magnetic field and the gap completes the qualitative understanding of type-I and type-II superconductors. In any event, BCS theory provides a microscopic quantitative account for these effects.

4.6 Microscopic origin for the dependence of transport behavior on temperature

Figures 4.8, 4.10, and 4.26 sketch the qualitative behavior of electric and thermal conductivities in metals and semiconductors as a function of temperature. The aim of this section is to provide convincing evidence that this behavior is related to scattering mechanisms where the carriers interact with phonons and impurities, the electron-electron scattering providing negligible contribution. The two types of processes are assumed to act in independent manner, as described in (4.60).

Considering that τ^{-1} measures the probability per unit time that a particle be scattered:

Concept

At a phenomenological level, one can expect that $\tau \propto 1/n_{scatt}$, with n_{scatt} the number of scatterers, either phonons or impurities, that are present in the material. In this view, the $\tau(T)$ dependence is dictated by the $n_{scatt}(T)$ dependence.

This section investigates the different mechanisms that may originate scattering of electrons or holes, identifying their dependence on temperature from the T -dependence of the number of scatterers in accord to the above rule of thumb. A bold list includes presence of impurities and whatever topological defect, interaction with phonons and with other electrons or holes. From now on, the term carriers is used to refer to both.

4.6.1 Scattering with impurities or defects

The presence of impurities or topological defects breaks the translational symmetry of the crystal and the Bloch theorem with it. In particular, the interaction of carriers with impurities or defects determines their transitions from an initial to a final state that in most cases belong to the same band. The average time τ_{im} lasting between two scattering events is expected to shorten as long as the concentration n_{im} of impurities or defects increases. In fact, the basic assumption here is that $\tau_{im} \propto n_{im}^{-1}$. If this assumption is valid, τ_{im} turns out to be independent of temperature due to the fact that n_{im} does.

Notice that defects occur also when the material surfaces are reconstructed or contaminated. The role of these defects is minor in a macroscopic crystal, but can become significant while the linear size of the sample squeezes: in fact, while the linear sample size becomes smaller than the carrier mean-free path.

4.6.2 Scattering with phonons

Use of analogous assumptions in the case of scattering with phonons would predict that the scattering time be $\tau_{ph} \propto n_{ph}^{-1}$, with n_{ph} the number of phonons present in the material. The latter depends on temperature: in particular, as already found in Chap. 3 one has $n_{ph} \propto T$ for $T \gg \Theta_D$ and $n_{ph} \propto T^3$ for $T \ll \Theta_D$ on the scale of the Debye temperature. Therefore, at high temperatures $T \gg \Theta_D$ on the scale of Θ_D one has $\tau_{ph}(T) \propto T^{-1}$, and at low temperatures $T \ll \Theta_D$ one gets $\tau_{ph}(T) \propto T^{-3}$.

Quantum mechanics provides additional insight on the electron-phonon scattering mechanism. In Chap. 3 the probability per unit time has indeed been calculated

to first order in perturbation theory, that free neutrons undergo a scattering event by interaction with lattice vibrations. There, neutrons were assumed to have initially unperturbed free-particle like wavefunctions, the scattered wavefunctions being determined by powers expansion of the unknown neutron potential in the nuclei field with respect to the ion displacements. In the case of electrons (or holes), one can repeat the same calculation by closely following that performed for neutrons, with the obvious difference that now the unperturbed wavefunctions are the electron Bloch functions which can be expanded in plane waves. Assuming that the transition occurs within the same electron band, to first order in perturbation theory the following selection rules are to be satisfied for the scattering event to be allowed:

$$\begin{aligned} \mathbf{k} - \mathbf{k}' &= \mathbf{q} + \mathbf{G} \\ E_n(\mathbf{k}) - E_n(\mathbf{k}') &= \pm \hbar \omega_s(\mathbf{q}), \end{aligned} \quad (4.136)$$

where \mathbf{k}, \mathbf{k}' and $E_n(\mathbf{k}), E_n(\mathbf{k}')$ are the wavenumbers and energies, respectively, of the initial and final (primed) states. Also, \mathbf{q} and $\hbar \omega_s(\mathbf{q})$ are the wavenumber and energy of the s -band phonon that is either created or destroyed in the scattering process. This problem is solved in deeper detail in Chap. 5.

The treatment of scattering mechanisms against phonons is more complicated in nanoscopic systems: as already seen indeed, these are realized by dispersing nanostructures in a hosting material, thereby requiring some sort of statistical averaging of the properties over the many non-identical items.

4.6.3 Scattering with other charge carriers

When scattering with phonons and impurities, typical involved energies may lead carriers to undergo transitions between different bands. When scattering with other carriers instead, the energies involved are typically not sufficient and occurrence of interband transitions is unlikely. Therefore,

Concept

The case of completely filled bands is distinguished because no electron-driven scattering, either interband or intraband, is relevant at all.

Indeed, besides the fact that interband transitions events are unlikely, intraband transitions result to be forbidden by the Pauli exclusion principle, no vacant states being available into which the carriers may scatter into.

In the following, the case of a partially filled conduction or valence band is therefore considered. In this situation, electron-electron scattering might play a significant and complex role in scattering mechanisms. Simple estimates support the latter statement. Given the total scattering cross-section σ_u for electron-electron interactions, an electron with velocity v spans in time dt the volume $\sigma_u v dt$. The number of

scattering events results then to be $dN_u = n\sigma_u v dt$, in terms of the electron volume density n . Since $N_u = n\sigma_u v$ is the number of collisions per unit time, the average time τ_{el} elapsing between two subsequent collisions is $\tau_{el} \simeq 1/N_u \sim 1/(n\sigma_u v)$ and the corresponding mean-free path is $\ell = v\tau_{el} = 1/(n\sigma_u)$. The electron density is actually better replaced with an effective density of carriers that participate to the scattering process: as already seen in the Sommerfeld theory for metals and in the treatment of extrinsic semiconductors, this actually corresponds to a typical and realistic situation. In conclusion, one has $\tau_{el} \sim 1/(n_{eff}\sigma_u v)$ and $\ell \sim 1/(n_{eff}\sigma_u)$, the scattering time τ_{el} thus depending on the three governing parameters: n_{eff} , σ_u , and v .

As it will be discussed in detail in Chap. 5, in metals $\sigma_u \propto k_f^{-2}$ and therefore $\ell \simeq k_f^2/n_{eff}$. In addition, as discussed in the Sommerfeld theory for metals, $n_{eff} \simeq nk_b T/\varepsilon_f$. Therefore, the expressions $\ell \sim n^{-\frac{1}{3}}\varepsilon_f/(k_b T) \sim 10^{-4} - 10^{-3}$ cm and $\tau_{el} \sim 10^{-12}$ s are obtained, the numerical estimates being performed at room temperature.

The Pauli principle does not affect the system behavior when the Fermi energy is comparable to or smaller than $k_b T$: such a condition occurs e.g. for $n < 10^{18}$ cm $^{-3}$. This is the case of semiconductor materials that are either intrinsic or extrinsic with medium-to-high doping levels. For semiconductors, typical estimates of ℓ and τ_{el} at room temperature are in fact: $\ell \sim 10^{-4} - 10^{-5}$ cm and $\tau_{el} \sim 10^{-10} - 10^{-11}$ s.

In both cases,

Concept

Electron-electron scattering events are seen to be rare processes in both bulk metals and semiconductors. This conclusion may change in extreme conditions of e.g. reduced dimensionality or very strong coupling interactions.

In the following of this section, the temperature dependence of the transport properties of metals and semiconductors is analyzed. The analysis stems from the assumptions that $\tau_{scatt} \propto n_{scatt}^{-1}$ and that to first approximation only phonons and impurities significantly contribute to scattering, electron-electron scattering being neglected at the present discussion level. The analysis is actually reduced to calculate in each case the effective number of scatterers as a function of temperature.

4.6.4 Metals

4.6.4.1 Electric conductivity

The starting step is relation (4.50), with the use though of the effective mass. Since the electron density in metals is temperature independent, the only T -dependence comes about from $\tau(T)$. As already seen, for high temperatures $T \gg \Theta_D$ one has $\tau^{-1} \gg \tau_{imp}^{-1}$ and one finds that $\sigma \propto T^{-1}$. For vanishing temperatures $T \rightarrow 0$, $\tau_{imp}^{-1} \gg \tau^{-1}$ and $\sigma \propto T^0$. The comprehension of the behavior in the regime with

$T \ll \Theta_D$, where $\sigma \propto T^{-5}$, needs some additional careful considerations. It has been shown that one of the effects of applying an uniform electric field is that the Fermi sphere be rigidly shifted by a quantity $\delta k \ll k_f$ for not so intense electric field. Once the electric field is removed, the system clearly tries to go back to equilibrium by successive scattering events. The question is how this occurs at the different temperatures.

The selection rules (4.136) help in understanding this point, since they establish that phonons must exist to serve the purpose of providing electrons with right wavenumbers and energies satisfying the related conservation laws: in this case, a large k of the order of k_f and a tiny energy are needed. At large temperatures $T \gg \Theta_D$, also only one phonon may work to this purpose. At low temperatures $T \ll \Theta_D$, phonons with such characteristics are not available: thus, equilibrium is to be reached by a mechanism involving a large number of scattering events with as many acoustical phonons. The change of electron wavenumber is represented in Fig. 4.29 for a scattering with a phonon of energy $\hbar\omega \simeq \hbar c_s q \sim k_b T$. This is:

$$\begin{aligned} \Delta k_x &= k_f (1 - \cos \theta) \simeq \frac{k_f}{2} \theta^2 \\ &\simeq \frac{k_f q^2}{2 k_f^2} = \frac{\omega^2}{2 k_f c_s^2} \simeq \frac{(k_b T)^2}{2 k_f c_s^2 \hbar^2}, \end{aligned} \quad (4.137)$$

with c_s the speed of sound and θ the angle between the directions of the electron wavenumbers \mathbf{k} and \mathbf{k}' . Considering that $q \ll k_f$, tiny momenta and energies are actually exchanged: the scattering is quasi-elastic and $k \simeq k'$, the outgoing electrons slightly deviating from the incoming ones. For the electron to access an empty state, a change $\delta k \simeq 2k_f$ would be needed in its wavenumber, corresponding to the number of collisions

$$n_u \simeq \frac{2k_f}{\Delta k_x} = \frac{4k_f^2 c_s^2}{(k_b T)^2}. \quad (4.138)$$

The effective number of phonons that is needed to realize the transition is therefore $n_{eff} = n_{ph}/n_u \propto T^5$. Here it comes the $\sigma \propto T^{-5}$ scaling law, that explains the qualitative experimental behavior in Fig. 4.8.

4.6.4.2 Thermal conductivity

The starting step is (4.66). Using the results obtained so far, one reaches the following conclusions about the properties of thermal conductivity:

Properties

P1. For $T \gg \Theta_D$, $c_V \propto T$, $\tau \propto T^{-1}$ and thus $\chi_c \propto T^0$.

$$|\mathbf{q}| = 2k_F \sin(\theta/2)$$

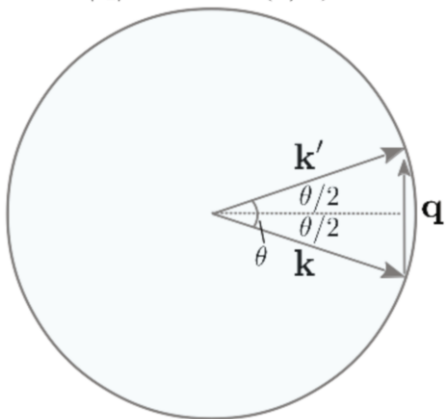


Fig. 4.29 Temperature-dependence of transport properties. Phonon scattering mechanism in metals at low temperatures. Sketch of the geometrical relationship between \mathbf{q} , \mathbf{k} , \mathbf{k}' , k_F and θ , useful to understand the $\sigma \propto T^{-5}$ qualitative behavior reported in Fig. 4.8

P2. For $T < \Theta_D$, $c_V \propto T$, $\tau \propto T^{-3}$ and thus $\chi_c \propto T^{-2}$.

P3. For $T \rightarrow 0$ when almost no phonons are around, $c_V \propto T$, $\tau \propto T^0$ predominantly driven by impurities and therefore temperature independent, and thus $\chi_c \propto T$.

These results completely explain the qualitative experimental behavior sketched in Fig. 4.10.

4.6.4.3 Thermoelectric effect

Expression (4.66) shows that the thermoelectric coefficient χ_t does not contain τ and is thus independent of collisional mechanisms. Therefore, the linear behavior predicted by the Fermi-Dirac statistics within the Sommerfeld theory of metals explains the experimental findings.

4.6.5 Semiconductors

4.6.5.1 Electric conductivity

Electric current in semiconductors is determined by both electrons in conduction band and holes in valence band. In Sec. 4.2.5 it has been inferred that electrons in the conduction band and holes in the valence band contribute to the current with

terms of the same sign. As such, it is convenient to introduce an electron and a hole conductivity.

The following notations are useful when describing the currents:

$$\begin{aligned}\sigma_n &= n_c \mu_n, \mu_n = \frac{e^2 \tau_n}{m_c} \\ \sigma_p &= p_v \mu_p, \mu_p = \frac{e^2 \tau_p}{m_v},\end{aligned}\quad (4.139)$$

which define two new and useful quantities, the electron μ_n and hole μ_p mobility. Expression (4.139) shows that the mobilities are dictated only by the effective masses m_c and m_v , along with the scattering times τ_n and τ_p . The latter, in turn, depend on the number of phonons and of impurities through the relations $\tau_n, \tau_p \propto n_{ph}^{-1}$ and $\tau_n, \tau_p \propto n_{im}^{-1}$ respectively. As a consequence, mobilities μ_n and μ_p depend on T via τ_n and τ_p . The temperature dependence of σ_n and σ_p instead, requires also the knowledge of the $n_c(T)$ and $p_v(T)$ functions.

Two limiting regimes might occur. The number of electrons or holes can be so large that the Fermi statistics is to be used: in this case the temperature behavior of μ_n and μ_p is similar to that found for metals. The number of electrons or holes can instead be so small that the Boltzmann statistics be used, as done e.g. in (4.113) and (4.114). In this case the temperature behavior of electric conductivity is as follows.

Properties

P1. If collisions involve acoustical phonons, as it occurs for $T \ll \Theta_D$, then $\mu_n \propto T^{-3}$. Assuming that $n_c, p_v \propto T^{\frac{3}{2}} e^{-a/T}$, one then finds $\sigma_n, \sigma_p \propto T^{-3/2} e^{-a/T}$.

P2. If collisions involve ionized impurities, then $\mu_n, \mu_p \propto T^0$. Therefore $\sigma_n, \sigma_p \propto T^{3/2} e^{-a/T}$.

P3. If instead the collisions involve optical phonons, as it occurs for $T \gg \Theta_D$, then $\mu_n, \mu_p \propto T^{-3/2}$, as detailed just below. Thus, $\sigma_n, \sigma_p \propto T^0 e^{-a/T}$.

Indeed, the number of optical phonons n_{op} is

$$n_{op} \propto \int_{\omega_0}^{\infty} \frac{\sqrt{\omega - \omega_0}}{e^{\frac{\hbar\omega}{k_b T}} - 1} d\omega = \left(\frac{k_b T}{\hbar} \right)^{\frac{3}{2}} \int_{x_0}^{\infty} dx \frac{\sqrt{x - x_0}}{e^x - 1} \quad (4.140)$$

where $x_0 = \hbar\omega_0/(k_b T)$. The leading contribution $n_{op} \propto T^{3/2}$ is thus found, modulated by the temperature-dependence of the integral.

Notice that the manner in which impurities are positioned inside the material can significantly affect the electric conductivity of charge carriers. Consider for example an n -doped semiconductor, whose electric conductivity should vanish as $T \rightarrow 0$. This prediction is experimentally verified if the average distance between impurities is larger than the average size of the electronic bound state. If the impurities separation is comparable or smaller instead, transitions may occur corresponding to electrons that migrate from one impurity to another: this type of mechanism brings about a new conductivity term that is expected to have the same characteristic temperature dependence as tunneling phenomena do.

4.6.5.2 Hall effect

The Hall effect in semiconductor structures is complicated by the presence of two different types of charge carriers, electrons and holes. Both carriers have positive effective mass and opposite charge, so that they move in opposite directions under the action of an electric field. If a magnetic field is also present, as in the Hall effect, the two species of carriers tend to accumulate on the same surface side of the sample.

For only one carrier species, the relationship between the current and the electric and magnetic fields is given by (4.77), where $\sigma = ne^2\tau/m$ and the conductivity tensor components given by:

$$\begin{aligned}\sigma_{xx} = \sigma_{yy} &= \frac{\sigma}{1 + (\omega_c\tau)^2} \\ \sigma_{xy} = -\sigma_{yx} &= -\frac{\sigma\omega_c\tau}{1 + (\omega_c\tau)^2}.\end{aligned}\quad (4.141)$$

This result can be extended to the presence of two species, that is:

$$\begin{aligned}\sigma_{xx}^{(1)} = \sigma_{yy}^{(1)} &= \frac{\sigma_1}{1 + (\omega_{1c}\tau_1)^2}, \quad \sigma_{xy}^{(1)} = -\sigma_{yx}^{(1)} = -\frac{\sigma_1\omega_{1c}\tau_1}{1 + (\omega_{1c}\tau_1)^2} \\ \sigma_{xx}^{(2)} = \sigma_{yy}^{(2)} &= \frac{\sigma_2}{1 + (\omega_{2c}\tau_2)^2}, \quad \sigma_{xy}^{(2)} = -\sigma_{yx}^{(2)} = \frac{\sigma_2\omega_{2c}\tau_2}{1 + (\omega_{2c}\tau_2)^2}\end{aligned}\quad (4.142)$$

The meaning of the symbols is obvious. Defining $\hat{\sigma}_1$ and $\hat{\sigma}_2$ as the conductivity tensors for the two carrier species, one finds for the total current the result:

$$\mathbf{j} = (\hat{\sigma}_1 + \hat{\sigma}_2)\mathbf{E} = \sigma\mathbf{E}.\quad (4.143)$$

Appendix 4.14 develops in detail the calculations to determine the Hall coefficient. In some cases the expressions are significantly simplified as summarized in the following.

Properties

P1. Case with $\omega_{1c}\tau_1, \omega_{2c}\tau_2 \ll 1$. One finds

$$R_H = \frac{1}{ec} \frac{n_2\mu_2^2 - n_1\mu_1^2}{(n_1\mu_1 + n_2\mu_2)^2}. \quad (4.144)$$

This formula further simplifies if $n_1 = n_2 = n$ into the following:

$$R_H = \frac{1}{ec} \frac{\mu_2 - \mu_1}{\mu_2 + \mu_1}, \quad (4.145)$$

whereas if $\mu_1 = \mu_2 = \mu$ one has

$$R_H = \frac{1}{ec} \frac{n_2 - n_1}{(n_1 + n_2)^2}. \quad (4.146)$$

P2. Case with $\omega_{1c}\tau_1, \omega_{2c}\tau_2 \gg 1$. One obtains

$$R_H = \frac{1}{ec} \frac{1}{n_2 - n_1}. \quad (4.147)$$

P3. Notice that in all the cases above, R_H can be either positive or negative, depending on the relative concentration of the two charge carrier species.

In particular, (4.144), (4.145) and (4.146) yield $R_H = 0$. Thus, one has $E_y = 0$ if $n_2\mu_2^2 = n_1\mu_1^2$, $\mu_2 = \mu_1$ and $n_2 = n_1$. From (4.147), one also sees that $R_H \rightarrow \infty$ when $n_2 = n_1$, so that $j_x = 0$.

4.6.5.3 Contribution of lattice vibrations to the thermal conductivity

The thermal conductivity is as well determined by both carrier species, electrons and holes. The thermal conductivity can therefore be calculated from (4.66), after inserting the temperature dependence of the density of electrons and holes. Since the thermal conductivity drops with the density of free carriers, it may become comparable with the contribution of lattice vibrations. Expression (4.66), originally found for electrons, is still valid and is used if the energy carriers are phonons. To this aim, the various terms in

$$\chi_c = \frac{v^2}{3} \tau n c_V \quad (4.148)$$

are just to be reinterpreted. To do this it is convenient to start from (4.63) for the energy current density, written in the form

$$j_e = \frac{1}{2} v_x [U(x - l_x) - U(x + l_x)].$$

Here, v_x is the phonons average velocity in the x -direction, assumed to be independent of T , and $U(x - l_x)$ and $U(x + l_x)$ the vibrational energy density in the points where the phonon had the last interaction. To first order in l_x the previous equation becomes

$$j_e = -\frac{v^2}{3}\tau \frac{\partial U}{\partial T} \frac{\partial T}{\partial x},$$

since $l_x = v_x \tau$, $v_x^2 = v^2/3$ and $\partial U/\partial T$ is the specific heat per unit volume. This formula is coincident with (4.148) if v^2 is the phonon average quadratic velocity, $nc_V = \partial U/\partial T$ and τ the average time between two successive collisions of a phonon with impurities or, in the presence of anharmonic effects, other phonons. A temperature gradient determines a thermal energy flux from regions at higher towards regions at lower temperatures. Thus, the behavior of χ_c with temperature is known if the T -dependence is known for c_V , v^2 and τ . In particular, the relevant findings are summarized below.

First, the speed of phonons v is dictated by the interaction strength of ion-ion coupling, the latter being only slightly dependent on temperature: as long as $T \ll T_D$, the constant above reduces to the elastic coefficients. Second, $\partial U/\partial T \propto T^3$ for lower and constant for higher temperature. Third, τ is dictated by the prevailing scattering mechanism. In particular:

- (a) When no impurities are present and for $T \rightarrow 0$, only a very few phonons are around: phonon-phonon scattering is negligible and the mean-free path increases up to reach the linear sample size. The smaller is the sample, the sooner this occurs.
- (b) When no impurities are present and for higher temperatures $T \ll \Theta_D$, phonon-phonon scattering becomes relevant and anharmonic terms may change the number and the frequency of the involved phonons, provided that the following conditions are satisfied:

$$\begin{aligned} \sum_{\mathbf{k},s} \mathbf{k} n_s(\mathbf{k}) &= \sum_{\mathbf{k},s} \mathbf{k} n'_s(\mathbf{k}) + \mathbf{G} \\ \sum_{\mathbf{k},s} n_s(\mathbf{k}) \omega_s(\mathbf{k}) &= \sum_{\mathbf{k},s} n'_s(\mathbf{k}) \omega_s(\mathbf{k}), \end{aligned} \quad (4.149)$$

in terms of the reciprocal lattice vector \mathbf{G} and of the phonon distributions $n_s(\mathbf{k})$ before and $n'_s(\mathbf{k})$ after the scattering event. Involving more than one phonon, these relations are of course based on scattering processes at next-to-first order. For normal processes, i.e. with $\mathbf{G} = 0$, wavenumber and total energy of the phonon flux remain unchanged, even in the presence of a temperature gradient ∇T . Only Umklapp processes, i.e. with $\mathbf{G} \neq 0$, may change wavenumber and energy of the phonon distributions. In this case, the total number of phonons is $n_t \propto T^3$ for $T < \Theta_D$, already seen to be acoustic-like and with small wavenumber. This number of phonons is to be effectively reduced by a factor $e^{-\Theta_D/T}$, since the number of phonons in an acoustical

branch with wavevector \mathbf{k} far from $k = 0$ is $n_s(k) = \left[e^{\hbar \omega_s(\mathbf{k})/(k_b T)} - 1 \right]^{-1} \simeq e^{-\Theta_D/T}$.

In addition, after each collision the phonon modifies its wavenumber by the amount $\Delta k_x = G(1 - \cos \varphi) \simeq G\varphi^2/2 \simeq G(\hbar \omega)^2/(2(c_s q_D)^2) \simeq G(\hbar k_b T)^2/(2(c_s q_D)^2)$. Thus, a number of collisions n_u such that $n_u \Delta k_x = G$ implies $n_u \propto T^{-2}$. The effective num-

ber of collisions is therefore proportional to $n_i/n_u \propto T^5 e^{-\Theta_D/T}$. All the above results gathered together lead to $\tau \propto T^{-5} e^{\Theta_D/T}$;

(c) When impurities are present, these contribute to τ in a manner that depends on their concentration and does not depend on temperature.

The above results are further summarized in Fig. 4.30 and in the following properties:

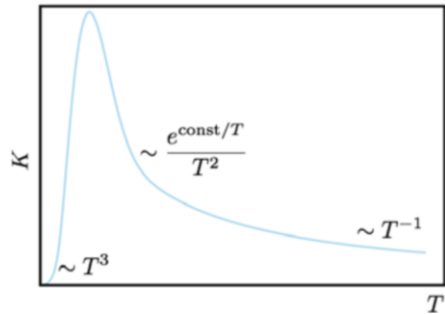
Properties

P1. $\chi_c \propto T^3$ as $T \rightarrow 0$ and/or small sample sizes. Phonon-phonon scattering is highly depressed.

P2. $\chi_c \propto T^{-2} e^{\Theta_D/T}$ for $T \ll \Theta_D$

P3. $\chi_c \propto T^{-1}$ for $T \gg \Theta_D$ because c_V is constant, the number of phonons is $\propto T$ and the scattering processes are dominated by Umklapp mechanisms.

Fig. 4.30 Temperature-dependence of transport properties. Thermal conductivity of semiconductors. Sketch of qualitative behavior of the phonon contribution

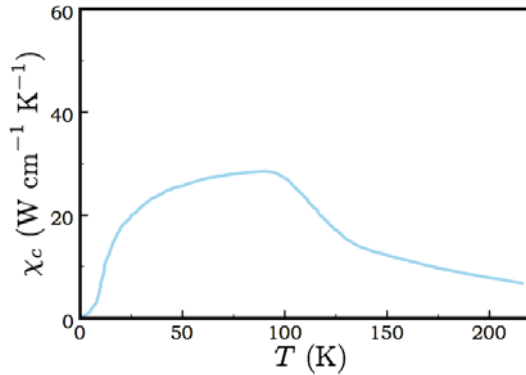


The theoretical prediction sketched in Fig. 4.30 is consistent with the experimental findings, as it can be verified by looking at Fig. 4.31, where the measured thermal conductivity of diamond is displayed [22]. The electronic contribution being negligible in an insulator, comparison between the two figures supports the theoretical predictions on the phonon contribution to χ_c in semiconductor materials.

4.7 Devices based on $p - n$ junctions

Special attention is now dedicated to the transport behavior of devices made of semiconductor-based $p - n$ junctions. The engineering of $p - n$ junctions and of semiconductor-based transistors has indeed been a crucial step in the development

Fig. 4.31 Temperature-dependence of transport properties. Thermal conductivity of semiconductors. Measured thermal conductivity of diamond as a function of temperature [22], supporting the theoretical prediction of Fig. 4.30 for the phonon contribution to χ_c in semiconductors. The electronic contribution is negligible because the material is insulating



of miniaturized electronic devices. It is well known that earlier diodes were made as vacuum cruets terminated by cathodic and anodic circuit ends, which a voltage is applied to. Electrons emitted by the cathodic end can be either accelerated or blocked according to the applied voltage, so that the device may work as a current rectifier. Of course, in the cruet the emitted electrons originate a space charge that strongly contributes to the process of favoring or blocking the current flow. Insertion of a grid in the cruet, whose potential can be changed from the outside, makes a triod which may work as a transistor: this is in fact designed to have a control of current flow by external governing knobs.

$p-n$ junctions represent the evolution of those electronic diodes, and two $p-n$ junctions in either one of the configurations $p-n-p$ or $n-p-n$ work as a transistor. The advantages of $p-n$ and $p-n-p$ or $n-p-n$ junctions over their valve-like ancestors are in their extremely reduced size and thus low electric powers involved. Typical voltages in diodes or triodes for example were of the order of ~ 10 V that, along with the macroscopic device size, determined high rates of energy dissipation by Joule effect. Semiconductor-based devices have the advantage of working with lower voltages and being a way much smaller in size.

4.7.0.4 Properties at thermodynamical equilibrium

The design of $p-n$ junctions requires technologies able to dope different regions of the same semiconductor sample in differential manners. A semiconductor doped with this characteristics is named inhomogeneous. In order to catch the relevant physics with simplest equations, the following model is considered to describe the spatial distribution of impurities in an inhomogeneous semiconductor designed to have p -doped and n -doped regions:

$$\begin{aligned}
 N_d(x) &= \begin{cases} N_d & \text{for } x > 0 \\ 0 & \text{for } x < 0 \end{cases} \\
 N_a(x) &= \begin{cases} N_a & \text{for } x < 0 \\ 0 & \text{for } x > 0 \end{cases} ,
 \end{aligned} \tag{4.150}$$

where $N_d(x)$ and $N_a(x)$ represent the density of donors and acceptors, respectively. If the two portions of the semiconductor were far apart, the value of the chemical potential in the donor part would be higher than that in the acceptor part, according to the discussion in Sec. 4.4.4. If the two parts are regions of the same sample material, at thermodynamical equilibrium the chemical potential is to be by definition the same in both portions: this can be the case only if electrons (holes) migrate from the donor (acceptor) region to the acceptor (donor) one. As a result, an electric field sets in proximity of the junction between the two regions: this is the origin of the given $p-n$ junction name. A schematic picture of the migration process and the consequent built-in electric field and potential is given in Fig. 4.32. In essence, when a n -type and a p -type semiconductor are brought together, conduction electrons shift from the n -type to the p -type semiconductor slab, creating an electrically polarized depletion region at the junction with an excess of positive charges in the n -type and of negative charges in the p -type regions. The electric field generated by the charge-depleted profile across the junction hampers any additional flow of majority carriers, that is electrons in n -region and holes in p -region, towards the region where they are minority carriers, that is electrons in p -region and holes in n -region, respectively. The same electric field favors instead the opposite flow, that is of carriers from region where they are minority-like towards region where they are majority-like. At thermodynamical equilibrium, both electrons and holes flows are to be separately null, in order to prevent a built-up of charge carriers in the two regions.

Study of the junction properties thus requires first of all the determination of the electric potential function ϕ that is established at thermodynamical equilibrium as the electric field sets in. To this aim, the two following simplifying operational conditions are adopted. First, $n_c \ll N_c$ and $p_v \ll P_v$ so that (4.113) and (4.114) can be used. Second, ϕ is assumed to be a slowly-varying function of space position x so that the semiclassical approximation can be used. Third, the migration process is considered to occur only in proximity of the junction, so that at equilibrium all donors and acceptors are ionized. As a result, one may write:

$$\begin{aligned}
 n_c(x) &= N_c(T) e^{-\frac{\varepsilon_c - e\phi(x) - \mu}{k_b T}} \\
 p_v(x) &= P_v(T) e^{-\frac{\mu - \varepsilon_v + e\phi(x)}{k_b T}} .
 \end{aligned} \tag{4.151}$$

The functions $n_c(x)$ and $p_v(x)$ satisfy the boundary conditions

$$\begin{aligned}
 n_c(x)_{x \rightarrow \infty} &= N_d = N_c(T) e^{-\frac{\varepsilon_c - e\phi(\infty) - \mu}{k_b T}} \\
 p_v(x)_{x \rightarrow -\infty} &= N_a = P_v(T) e^{-\frac{\mu - \varepsilon_v + e\phi(-\infty)}{k_b T}} .
 \end{aligned} \tag{4.152}$$

From (4.152) it is easily found that

$$e\Delta\phi = e[\phi(\infty) - \phi(-\infty)] = E_g + k_bT \ln \left(\frac{N_d N_a}{N_c P_v} \right). \quad (4.153)$$

That is, a step occurs in the electric potential function, as sketched in the lowest panel of Fig. 4.32. The step amounts to $|e\Delta\phi| \sim E_g$ and is modulated by a temperature-dependent term. The expressions (4.151) can also be cast in the simpler form

$$\begin{aligned} n_c(x) &= N_d e^{-\frac{e(\phi(\infty) - e\phi(x))}{k_bT}} \\ p_v(x) &= N_a e^{-\frac{e(\phi(x) - e\phi(\infty))}{k_bT}}. \end{aligned} \quad (4.154)$$

The size of the region $-d_p < x < d_n$ where the potential ϕ changes by $\Delta\phi$ is a relevant quantity to be determined, as sketched in the top panel of Fig. 4.32. A few considerations help here. First, since $k_bT \ll e|\Delta\phi|$, n_c and p_v can be neglected inside this region $-d_p < x < d_n$. In essence, no charge carriers are present in the junction region with size $d_p + d_n$, the latter being in fact named depletion region. Finally, $N_d \neq 0$ for $x > 0$ and $N_a \neq 0$ for $x < 0$. Considering all of the above and using the boundary conditions for n_c and p_v , the charge density can be expressed as

$$\rho(x) = e[N_d(x) - N_a(x) - n_c(x) + p_v(x)] \quad (4.155)$$

with

$$\begin{aligned} \rho(x) &= 0 & \text{for } x > d_n \\ \rho(x) &= eN_d & \text{for } 0 < x < d_n \\ \rho(x) &= -eN_a & \text{for } -d_p < x < 0 \\ \rho(x) &= 0 & \text{for } x < -d_p. \end{aligned} \quad (4.156)$$

The Poisson equation for $\phi(x)$ is therefore:

$$\frac{d^2\phi(x)}{dx^2} = -\frac{4\pi}{\epsilon_0}\rho(x), \quad (4.157)$$

with ϵ_0 is the static dielectric constant of the material. The equation, with the boundary conditions $\phi(x)_{x \rightarrow \infty} = \phi(\infty)$ and $\phi(x)_{x \rightarrow -\infty} = \phi(-\infty)$, is easily solved requiring conditions of continuity for $\phi(x)$ and its first derivative. One then finds:

$$\begin{aligned} \phi &= \phi(\infty) & \text{for } x > d_n \\ \phi &= \phi(\infty) - \frac{2\pi e N_d}{\epsilon_0} (x - d_n)^2 & \text{for } 0 < x < d_n \\ \phi &= \phi(-\infty) + \frac{2\pi e N_a}{\epsilon_0} (x + d_p)^2 & \text{for } -d_p < x < 0 \\ \phi &= \phi(-\infty) & \text{for } x < -d_p. \end{aligned} \quad (4.158)$$

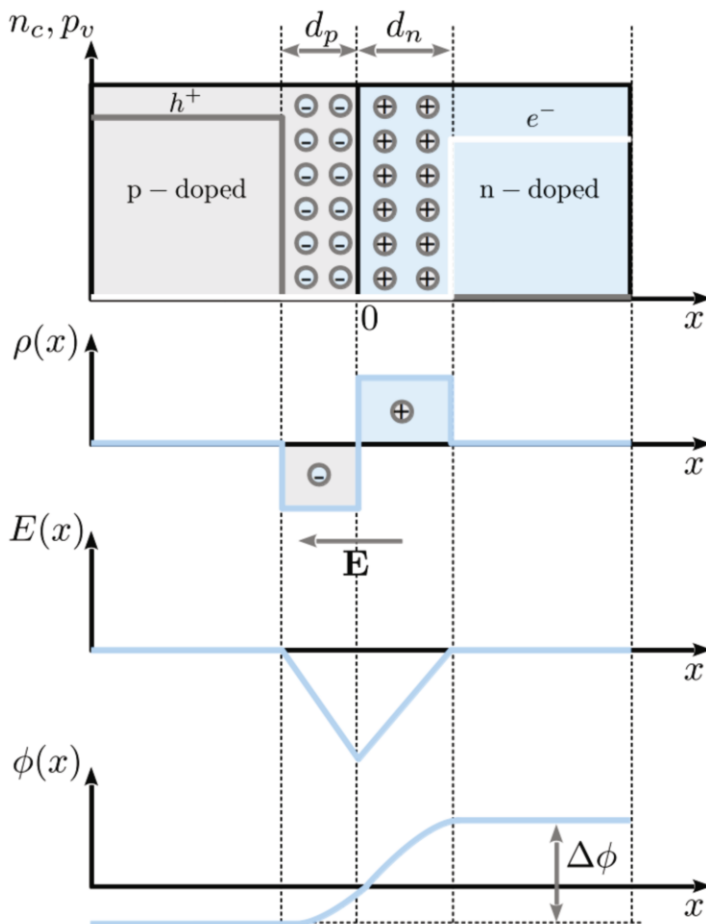


Fig. 4.32 How a $p-n$ junction works. The characterizing quantities discussed in the text are sketched: space charge layers in the p and n region; their relative widths d_p and d_n , respectively; the free carrier densities n_c for electrons and p_v for holes; the charge distribution of the ionized impurities $\rho(x)$; the built-in electric field \mathbf{E} ; the corresponding electrical potential $\phi(x)$; the built-in potential difference $\Delta\phi$

The function $\phi(x)$ in (4.158) satisfies the continuity condition except than at the origin. Here, the conditions $\phi(0^+) = \phi(0^-)$ and $d\phi(x)/dx|_{0^+} = d\phi(x)/dx|_{0^-}$ are to be imposed, leading to:

$$\begin{aligned} \Delta\phi &= \frac{2\pi e}{\epsilon_0} (N_d d_n^2 + N_a d_p^2) \\ N_d d_n &= N_a d_p. \end{aligned} \quad (4.159)$$

From (4.159), the lengths $d_{n,p}$ can be extracted along with the size of the depletion region:

$$d_{n,p} = \sqrt{\frac{(N_a/N_d)^{\pm 1} \epsilon_0 \Delta \phi}{(N_d + N_a) 2\pi e}}, \quad (4.160)$$

where the plus and minus signs refer to d_n and d_p , respectively.

The chemical potential can now be calculated from (4.152):

$$\mu + e\phi(-\infty) = \epsilon_V + k_b T \ln \frac{P_V(T)}{N_a}. \quad (4.161)$$

The arbitrary constant $e\phi(-\infty)$ can be set to zero if the reference potential is taken to be that of the p -doped portion. Realistic values of d_n and d_p range from 10^{-6} to 10^{-4} cm, supporting a posteriori the use of the semiclassical approximation.

Concept

Thermodynamical equilibrium conditions in a $p-n$ junction manifest with the appearance of a charge depletion and of non-uniform electric potential profiles across the junction, ensuring that the chemical potential be the same within the whole device.

When an external potential V is applied between the p and n regions, equilibrium conditions are altered. If electrons are added to the n -type semiconductor and removed from the p -type, that is in the presence of a *forward bias voltage* V , the positive (negative) charge excess in the n -(p -)type semiconductor is neutralized and the size of the depletion region decreases: current can therefore flow across the junction. If electrons are instead removed from the n -type and added to the p -type semiconductor as in a so-called *reverse bias voltage*, the depletion region is further enlarged in spatial size, becoming so thick that no current may flow across the junction. The $p-n$ junction therefore works as a rectifier, depending on the bias being forward or backward.

Quantitatively, the $p-n$ junction works as follows. A conventional sign is from here on used, establishing that V be positive (negative) if the p -region is at higher (lower) potential with respect to the n -region. The potential barrier $\Delta\phi$ in the new situation thus becomes:

$$\Delta\phi_V = \Delta\phi - V.$$

Therefore, the potential step $\Delta\phi$ is reduced (enhanced) if $V > 0$ ($V < 0$). The size of the depletion region changes as well, since it depends on $\Delta\phi$:

$$d_{n,p}(V) = d_{n,p}(0) \left(1 - \frac{V}{\Delta\phi}\right)^{\frac{1}{2}}. \quad (4.162)$$

Thus, one has $d_{n,p}(V) < d_{n,p}(0)$ ($d_{n,p}(V) > d_{n,p}(0)$) if $V > 0$ ($V < 0$).

The lowering of the potential barrier when $V > 0$ affects the number of majority carriers and leaves to first approximation the number of minority carriers un-

changed. In fact, the number of majority carriers migrating per unit time into the region where they are minority-like increases. As a result, an electric-current flow sets in, composed of electrons and holes. The current of minority carriers is basically unaltered, since it mainly depends on the mechanisms taking place outside the depletion region and leading to generation of minority carriers. The current is therefore

$$j_e = e[J_e(V) - J_e^{gen}],$$

where the first term accounts for majority and second for minority carriers. Since $j_e = 0$ when $V = 0$, that is $J_e(V = 0) = J_e^{gen}$, it follows that the number of electrons that are able to overcome the potential barrier exponentially increases (decreases) with decreasing (increasing) V . Therefore, the relation for the current can be cast into the more expressive form

$$j_e = eJ_e^{gen} \left(e^{\frac{eV}{k_b T}} - 1 \right). \quad (4.163)$$

The hole current can be treated along the same reasoning lines as for the electron current.

Getting all together, the total current can be expressed as

$$j = e(J_e^{gen} + J_h^{gen}) \left(e^{\frac{eV}{k_b T}} - 1 \right). \quad (4.164)$$

This relation represents the characteristic function of the $p-n$ junction governing the changes of current after changes of the external potential V . The $j-V$ characteristic behavior is displayed in Fig. 4.33, along with the schematic view of the

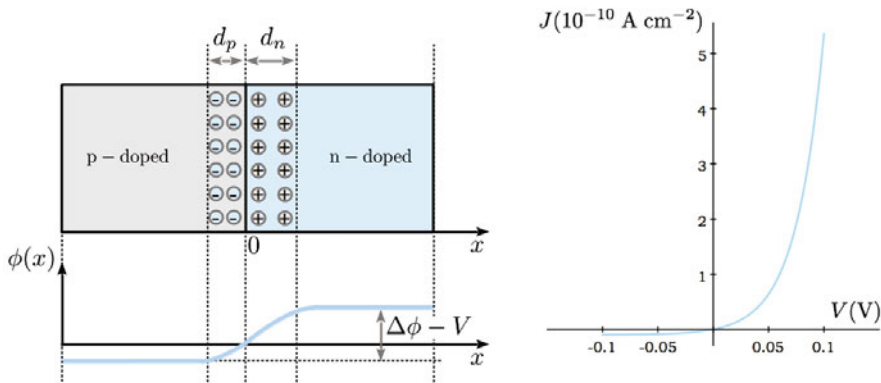


Fig. 4.33 Transport behavior of $p-n$ junctions. (a) Scheme of the electric potential profile across the junction in the case with a positive $V > 0$ applied external potential that lowers the barrier between n - and p -regions. (b) The resulting current-voltage characteristics, showing that the $p-n$ junction works as a current rectifier (see text)

potential profile across the junction. When $V < 0$, $j \rightarrow 0$. On the other hand, for $V > 0$ j exponentially increases and very rapidly reaches significant values, due to the fact that the condition $eV/(k_bT) \gg 1$ can be met even with tiny potentials: this is a relevant advantage with respect to older non-semiconductor based diodes. In essence,

Concept

In $p-n$ junctions, the current is exponentially enhanced ($V > 0$) or depressed ($V < 0$) after changes of the external potential V . The exponential decreasing behavior saturates to a vanishing value, whose size depends on the specific mechanisms generating the minority current. This current-voltage characteristics (4.164) makes the $p-n$ junction working as a current rectifier.

Appendix 4.15 completes the present treatment with the calculations of the minority currents J_e^{gen} and J_h^{gen} , providing additional insight on the saturation behavior of the current-voltage characteristics.

Quick Questions

Q27. In the back of the AC adapter of your laptop is stated that this is converting the AC from the outlet to a 19.5V DC voltage. The peak-to-peak AC voltage is 120 V. Guess the least electronic components that your adapter contains [1].

Answer - Current and voltage are to be rectified. So the power adapter must at least contain diodes. Diodes are not enough though. Since the outlet AC voltage has a peak-to-peak height of 120 V, a transformer is also needed inside the power adapter, that reduces this eight to about 19 V by electromagnetic induction from a primary coil with a larger number of windings to the secondary coil with a lower number of windings. Notice also that during the time in which the current drops to zero because of rectification, your laptop still needs a power supply: this could be ensured by the presence of a capacitor, acting as a charge and energy store that can be delivered whenever needed.

4.8 Electric conduction in nanostructures

4.8.1 Phenomenology

In both cases of metals and of semiconductors or insulators, the concept of local thermodynamic equilibrium has been used as a guiding condition to investigate electric, thermal and thermoelectric phenomena. In this collisional or hydrodynamic

regime the high rate of collisions rapidly brings the system to equilibrium. Application of this concept required the possibility of dividing the sample into small volumes, each at local thermodynamical equilibrium, that has to contain a large enough number of particles for the calculated statistical averages be meaningful.

On the other hand, it has already been seen that technological applications of semiconductor properties require sample miniaturization. On such small length scales, the local thermodynamical equilibrium concept breaks down. In the following, ways to circumvent this difficulty are discussed, while illustrating the working concept of nanostructure-based devices. Once the basic working concepts of semiconductor-based diodes are comprehended from Sec. 4.7 devoted to $p-n$ junctions, this section is devoted to nanostructure-based triodes, the transistors, that are a basic building block for electronic devices.

Consider a material placed between two metal plates connected to the nanostructure by means of contacts. In the following simplified model, the contacts are considered ideal, with no resistance. The device scheme is illustrated in Fig. 4.34. The two external contacts are made from the same metal: they work as source and drain terminals between which a voltage difference $V_D - V_S$ can be applied across the channel structure of nanometric size. When $V_D = V_G = 0$, the system is in thermodynamical equilibrium and the chemical potential μ is uniform: in fact, due to the tiny size of the semiconductor between the two plates, μ is the chemical potential typical of a metal. A third terminal, named gate, is at voltage V_G controls the current flowing between the source and the drain. An insulating layer separates the gate contact from the conducting channel. The insulating layer thickness is a trade-off between two different requirements: it has to be thick enough to offer very large resistance to the passage of current, and thin enough that a third applied potential may influence the electronic energy levels of the conducting channel. The electric current in the nanostructure is determined by the parameters $V_D - V_S$ and V_G . The first gives the energy range in which the electronic levels contribute to the current, in the device, the second moves these levels into or out the energy range, driving the system to an “on” (conduction) or “off” (insulating) state. The device above works as a transistor and is often referred to as a Metal-Oxide-Semiconductor Field-Effect Transistor or MOSFET.

The working principle of the MOSFET, detailed in the next section, can be understood as follows. When a bias voltage is applied between source and drain, the system is driven out of equilibrium, that is, different chemical potentials, say μ_S and μ_D are induced. By applying a positive (negative) bias voltage, the channel electronic level move down (up). The current flow intensity depends on the number of electronic states that, at the applied gate voltage, lie between μ_S and μ_D . This is illustrated in the lower part of Fig. 4.34.

As for the $p-n$ junction, the base information to be worked out here is the characteristic current-voltage function, that is the behavior of the current between source and drain after variations of V_D and V_G . Fig. 4.35 displays the typical current-voltage characteristics, as it is measured in experiments. In particular, Fig. 4.35(a) shows the current-gate voltage characteristic for selected values of V_D and Fig. 4.35(b) the current-drain voltage characteristic for selected values of V_G . The current-gate volt-

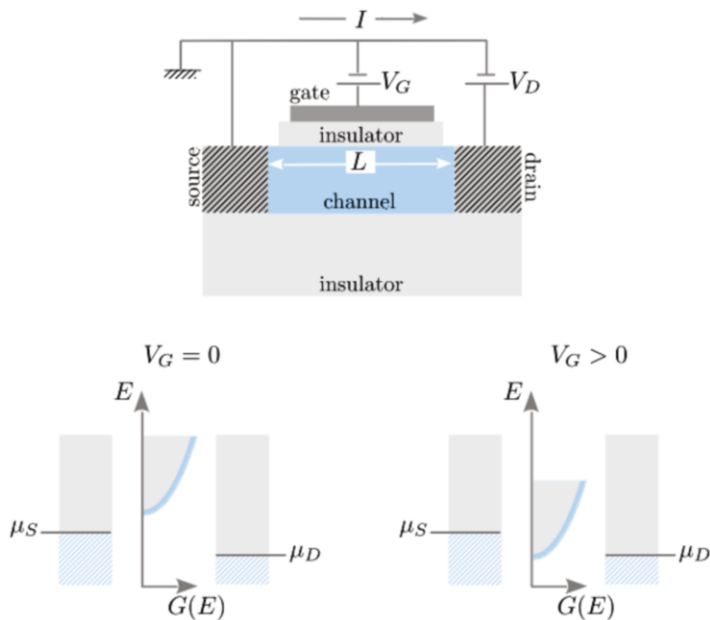


Fig. 4.34 Transport behavior of nanostructures. Top: scheme of a MOSFET as described in the main text, illustrating the characteristic three-terminal configuration and related voltages: source V_S , drain V_D , and gate V_G . Bottom: sketch of the density of states at source and drain terminals under two different gate driving, that are $V_G = 0$ on the left and $V_G > 0$ on the right

age curve is seen to be characterized by a threshold voltage V_T : below this applied value of gate voltage, no current is appreciably generated within the circuit. The value of V_T is observed to drop while V_D increases.

Quick Questions

Q28. n -channel and p -channel MOSFETs are often used in pairs to build

a so-called CMOS element: in this example, an n -channel MOSFET is connected to the negative terminal of the device power supply and the p -channel MOSFET to the positive one. Such a basic CMOS element works as a logic inverter, that is it outputs a 1 when the input is a 0 and viceversa: ones and zeros are represented by positive and negative charges, respectively, that arrive at the inverter input or get at its output. Guess how such an inverter might work [1].

Answer - If a negative charge arrives at the inverter input, that is a 0, only the p -channel MOSFET conducts current and the output is positive, that is a 1. The opposite occurs if a positive charge arrives at the input, this time activating only the n -channel MOSFET. Notice that inverters are basic logic elements to build any other logic element in circuitry, such as AND, NOT, NAND gates.

4.8.2 Conductance in nanostructures

The present section is aimed to comprehend the physics underlying the phenomenology illustrated in Sec. 4.8.1. Apply the voltage V_D between source and gate: the source μ_1 and gate μ_2 chemical potentials have different values, related by $\mu_2 - \mu_1 = -eV_D$. Indeed, the applied potential is slowly varying on the scale of the lattice parameter, and therefore the Fermi distribution functions of electrons in the source and drain metallic plates can still be used to characterize the system. Two different cases are conveniently distinguished, according to whether V_D is given positive or negative values. When $V_D > 0$ or $V_D < 0$, then $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$, respectively. Under the action of the external field, electrons are pushed from source to drain in the former case and from drain to source in latter.

If no electronic nanostructure states are available, which possess an energy in the interval between μ_1 and μ_2 , current flows can be established in the nanostructured channel only through tunnel mechanisms or more complicated higher-order processes involving more electronic nanostructure levels as intermediate states. If electronic states are instead available within the interval (μ_1, μ_2) , current flows between source and drain are strongly favored.

To make the treatment simple, the existence of only one such state is from now on assumed, that is singly occupied so that any spin-related mechanism can be neglected. Let ε be its energy. The further simplifying assumption is adopted for now, that to first approximation ε is not modified by drain and gate voltages. Though not fully realistic, this assumption helps catching a few more relevant concepts first, and it will be released later on in Sec. 4.8.3 within a more refined treatment. As any other states in the channel between source and drain, this state is influenced by the possibly applied V_D . The voltage V_G in turn, works to place the nanostructure electronic level at the desired energy.

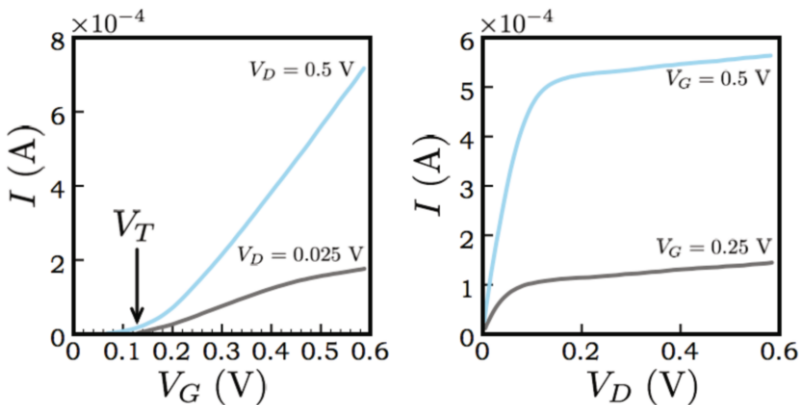


Fig. 4.35 Transport behavior of nanostructures. Measured current-voltage characteristics for the transistor type depicted in Fig. 4.34. (a) Current-gate voltage characteristic for selected values of V_D , showing also the typical threshold voltage V_T . (b) Current-drain voltage characteristic for selected values of V_G [4]

At any given V_G value, the current flowing in the channel from source to drain, say I_1 , or viceversa, say I_2 , can be expressed as:

$$\begin{aligned} I_1 &= \frac{e\gamma_1}{\hbar} (f_1 - N) \\ I_2 &= \frac{e\gamma_2}{\hbar} (f_2 - N). \end{aligned} \quad (4.165)$$

Here, N is the average occupation number of the nanostructure state, γ_1/\hbar and γ_2/\hbar are the probabilities per unit time that an electron travels from source to drain or viceversa, respectively. The functions $f_{1,2}$ are

$$f_1 = \frac{1}{1 + e^{\beta(\varepsilon - \mu_1)}}, \quad f_2 = \frac{1}{1 + e^{\beta(\varepsilon - \mu_2)}} \quad (4.166)$$

the Fermi distribution functions. Notice that they are evaluated at the energy ε of the nanostructure state: the passage of the electron from the metal plate to the nanostructure channel is simply guaranteed and driven by the overlap of its wavefunctions in the metal and in the channel state. The currents I_1 and I_2 are composed of two terms: as sketched in Fig. 4.36, electrons can in fact flow from the electrodes to the nanostructure as in the first term of (4.165)), and viceversa as in the second term of (4.165)).

The total current from source to drain is $I_t = I_1 + I_2$. Under stationary conditions, that is when $I_t = 0$ or $I_1 = -I_2$, one has

$$N = \frac{\gamma_1 f_1 + \gamma_2 f_2}{\gamma_1 + \gamma_2}. \quad (4.167)$$

Therefore, use of (4.167) yields the expressions for the currents I_1 and I_2 :

$$I_1 = -I_2 = I = \frac{e}{\hbar} \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} (f_1(\varepsilon) - f_2(\varepsilon)). \quad (4.168)$$

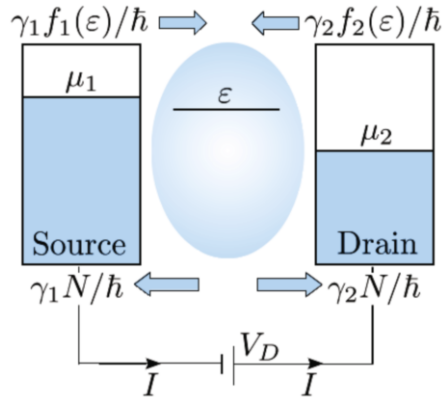


Fig. 4.36 Transport behavior of nanostructures. Properties of the transistor schematized in Fig 4.34. Scheme of chemical potentials and current flows established between source, drain, and channel in the case with one singly occupied level inside the nanostructured channel

Consider for example the case with $V_D > 0$, corresponding to $\mu_1 > \mu_2$. When $(\varepsilon - \mu_1)/(k_b T) \gg 1$, then $f_1(\varepsilon) \simeq f_2(\varepsilon) \simeq 0$ and therefore $I = 0$. When $(\mu_2 - \varepsilon)/(k_b T) \gg 1$, then $f_1(\varepsilon) \simeq f_2(\varepsilon) \simeq 1$ and once again $I = 0$. Thus, in order to have non-vanishing currents $I \neq 0$, $f_1(\varepsilon)$ and $f_2(\varepsilon)$ are to be significantly different. This may occur while $\mu_2 < \varepsilon < \mu_1$, that is if temperatures are low enough that the conditions $f_2(\varepsilon) \simeq 0$ and $f_1(\varepsilon) \simeq 1$ be realized. If this is the case, the current turns out to be

$$I = \frac{e}{\hbar} \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2}. \quad (4.169)$$

In the special case in which $\gamma_1 = \gamma_2 = \gamma$, one has the very simple result

$$I = \frac{e}{\hbar} \frac{\gamma}{2}. \quad (4.170)$$

Expressions (4.169) and (4.170) are especially interesting, as they represent the current of one single electron that depends only on the interaction mechanisms intrinsic to the material and governing $\gamma_{1,2}$: no dependence on the geometry of the sample manifests. This is a quite peculiar behavior indeed, contrasting that of ohmic-like samples: Ohm's law indeed states that the conductance G satisfies the relation $GV_D = I$, with $G = \sigma A/L$ and σ , A and L the sample conductivity, transverse section and length.

In fact, expression (4.170) does not depend on the drain voltage V_D and would imply an indefinitely increasing current I after increasing values of γ . Experiments show instead that the conductance cannot exceed the maximum value $G_0 = e^2/(2\pi\hbar)$. To understand and fix these flaws, a simplified model can be considered in which the given level inside the channel is parametrized as a resonant state with an energy width γ determined by the state lifetime. The energy of the state can thus be represented as distributed according to the following lorentzian shape:

$$D(E) = \frac{1}{2\pi} \frac{\gamma_1 + \gamma_2}{[E - \varepsilon]^2 + [(\gamma_1 + \gamma_2)/2]^2}, \quad (4.171)$$

that is typical of finite-lifetime energy states. The energy profile of the state is as more sharply peaked around the central value $E = \varepsilon$, as longer is its lifetime, i.e. as smaller is the quantity $\gamma = \gamma_1 + \gamma_2$. The curve is normalized so that $\int_{-\infty}^{+\infty} dE D(E) = 1$, in a way that the calculated average width $\int_{-\infty}^{+\infty} dE (E - \varepsilon)^2 D(E)$ is precisely γ . Notice that $D(E)$ tends to be represented by the Dirac delta-function $\delta(E - \varepsilon)$ in the extreme $\gamma \rightarrow 0$ limit.

Now, expression (4.170) misses a crucial observation: the larger is γ on the scale of the $\mu_1 - \mu_2$ difference, the more hampered is expected to be the current while flowing from source to drain. This remark can be qualitatively taken into account by reducing (4.170) by an amount of the order of $(\mu_1 - \mu_2)/(C\gamma)$, with C a numerical factor. In fact, the realistic expression for (4.170) is

$$I = \frac{e}{\hbar} \frac{\gamma}{2} \frac{\mu_1 - \mu_2}{C\gamma} = \frac{e^2}{\hbar} \frac{V_D}{2C}. \quad (4.172)$$

A more refined and quantitative analysis can be conducted starting from (4.168) and writing

$$I = \frac{e}{\hbar} \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} \int_{-\infty}^{\infty} dE D(E) (f_1(E) - f_2(E)). \quad (4.173)$$

Since

$$f_1(E) - f_2(E) = \begin{cases} 1 & \text{for } \mu_1 > E > \mu_2 \\ 0 & \text{otherwise,} \end{cases}$$

one has that

$$I = \frac{e}{\hbar} \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} \int_{\mu_2}^{\mu_1} dE D(E). \quad (4.174)$$

Under the simplifying assumption that $D(E)$ be approximately constant within the interval (μ_1, μ_2) , the current can be expressed as

$$I = \frac{e}{\hbar} \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} (\mu_1 - \mu_2) \frac{1}{2\pi} \frac{\gamma_1 + \gamma_2}{[\mu - \varepsilon]^2 + [(\gamma_1 + \gamma_2)/2]^2}. \quad (4.175)$$

Here, μ is the average between μ_1 and μ_2 : μ gets its maximum value when $\varepsilon = \mu$, that is when the chemical potential of the metal contact is resonant with the energy of the state in the nanostructure channel. At resonance one therefore obtains

$$I = \frac{e}{2\pi\hbar} \frac{4\gamma_1\gamma_2}{(\gamma_1 + \gamma_2)^2} (\mu_1 - \mu_2). \quad (4.176)$$

For $\gamma_1 = \gamma_2$ and $\mu_1 - \mu_2 = eV_D$, this yields

$$\frac{I}{V_D} = \frac{e^2}{2\pi\hbar}, \quad (4.177)$$

revealing that the constant C in the qualitative argument is precisely π . In conclusion:

Concept

Non-ohmic behavior characterizes transistor transport behavior. The source-to-drain conductance cannot exceed the value $G = I/V_D = e^2/(2\pi\hbar)$, that is dictated solely by fundamental constants related to the quantum particle, the sample geometry not playing any role.

Quick Questions

Q29. You are looking at the result of an experiment where the maximum conductance G of a nanostructure material has been measured to be $G_M = (1.55 \pm 0.02) \cdot 10^{-4} \Omega^{-1}$. Assuming that all the systematics are under control, what would you infer from this result?

Answer - Inserting the numbers for the fundamental constants, the data are compatible with a carrier charge that is twice that of the electron. An hypothesis could be inferred, according to which the system behave as if the carriers were composed of paired electrons (or holes) .

4.8.3 Current-drain and current-gate voltage characteristics

In the present section a few oversimplifying assumptions underlying the above treatment are released, leading to a more realistic analysis and deeper understanding of the I - V_D and I - V_G characteristics. Here, the modifications of ϵ due to the presence of drain and gate voltages are taken into account, along with the charge state of the junction.

The energy involved in the localization process of the electron inside the nanostructure can be calculated, schematizing the nanostructure to a dot and simulating the system to behave as the circuit schematized in Fig. 4.37. The charge $-e$ inside the nanostructure can be thought to be shared onto three capacitors located at source, gate, and drain respectively, so that $q_S + q_G + q_D = e$. The three capacitors

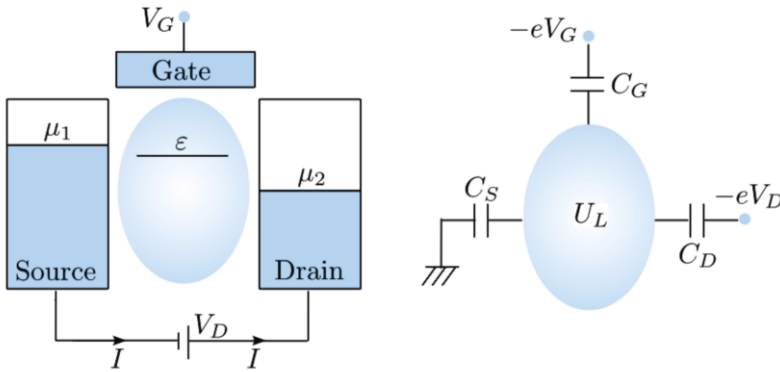


Fig. 4.37 Transport behavior of nanostructures. Properties of the transistor schematized in Fig. 4.34 [4]. Scheme illustrating the nanodot model used in the text, to help understanding the energy involved in the localization process of an electron inside the channel nanostructure. The system is simulated to behave as a circuit with three capacitance in parallel, located at source, drain, and gate

result to be in parallel, and therefore the nanostructure is at a potential V_0 such that $q_S = C_S V_0$, $q_G = C_G V_0$, and $q_D = C_D V_0$. Summing the three charges over, one obtains $e = (C_S + C_G + C_D) V_0 = C_E V_0$. The electrostatic energy of the system is thus

$$U_L = -q_G V_G - q_D V_D = -e V_0 \left(\frac{C_G}{C_E} + \frac{C_D}{C_E} \right). \quad (4.178)$$

If the charge in the nanometric zone changes by $-e\Delta N$, the potential V is modified according to

$$-e\Delta N = C_S V + C_G(V - V_G) + C_D(V - V_D),$$

or else

$$-e\Delta N = VC_E - (C_G V_G + C_D V_D).$$

One then finds for the potential energy $U = -eV$ the result

$$U = U_L + \frac{e^2 \Delta N}{C_E}, \quad (4.179)$$

where ΔN is calculated with respect to the reference situation with $N = 1$.

Definition

The quantity e^2/C_E represents the energy required to add one electron to the nanostructure and is therefore called charge energy.

In the presence of additional charges modifying the energy, the current is to be calculated through the following relations:

$$\begin{aligned} I &= \frac{e}{\hbar} \int D(E - U) dE \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} (f_1(E) - f_2(E)) \\ N &= \int D(E - U) dE \frac{\gamma_1 f_1(E) + \gamma_2 f_2(E)}{\gamma_1 + \gamma_2} \\ U &= U_L + \frac{e^2 \Delta N}{C_E}, \end{aligned} \quad (4.180)$$

with $D(E)$ given by (4.171). The calculations are to be performed in a self-consistent manner, since U depends on N . The resulting theoretical predictions agree, with the experimental data represented in Fig. 4.35 from Datta [4].

Quick Questions

Q30. Consider an electronic state inside a channel that is already occupied. Is a current flow from source to drain still possible?

Answer - The presence of an electron in the electronic state does not prevent a current flow, provided that the state can support the presence of two electrons and that the new value of its energy keeps lying in the interval (μ_1, μ_2) . Violation of either one of these two conditions would block any current flow, a situation that is indeed referred to as Coulomb blockade (as sketched in Fig. 4.38).

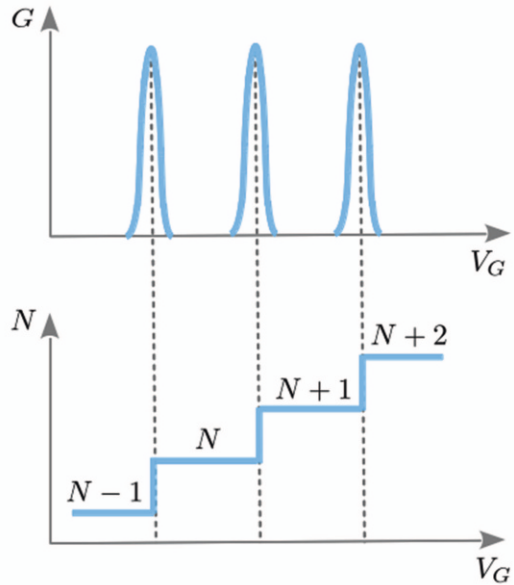


Fig. 4.38 Transport behavior of nanostructures. A schematic view of the conductance (top panel) and number of electrons (bottom panel) as a function of the gate voltage in a single-electron transistor. The Coulomb blockade effect shows up: the conductance is zero, unless a single charge tunnelling event takes place. This allows to add/remove electrons from the device one by one.

4.9 Experimental methods to determine transport behaviors

Experiments suited to determine the transport behavior of macroscopic sample materials, such as measurements of electric and thermal conductivity, are performed by means of standard methods: these are based on measurements of either electric or energy currents after a either voltage or a temperature gradient is applied, respectively.

The thermoelectric coefficient in metals and semiconductors is determined by measuring the electromotive force between the two opposite ends of a sample, which either a temperature gradient or an electric current have been applied to, at closed circuit. The corresponding measurements can be experimentally delicate, since a small effect is being looked for.

In the presence of a magnetic field, spectroscopy of the Landau states in the microwave range of the electromagnetic field can be used to obtain information about the cyclotron mass. Indeed, when the dipole matrix element is calculated in direction perpendicular to the magnetic field, only transitions between adjacent states can be observed because of the selection rule $\Delta n = 1$ that their principal quantum numbers n must satisfy. Measurements of Hall effect are as well performed by standard methods. Measurements related to quantum Hall effect instead require special care in sample preparation because they involve very low temperatures and quite strong magnetic fields: the oxide-semiconductor interface is to be accurately cleaned, in a way to effectively realize a two-dimensional electron layer.

The trend towards progressive and extreme miniaturization of the devices and of their electric working powers is systematically going on since several decades. The nanometric transistor discussed in the Sec. 4.8 is one example of such technological advances, though it is not even the most recent.

Finally, the determination of phonon transport properties requires a measurement of sound velocities for acoustical excitations: this is typically accomplished after sending short high-frequency pulses of longitudinal or transverse deformations in a macroscopic crystal by means of a piezoelectric transducer and measuring the travel time. Additional information can be extracted from the determination of phonon spectra, by means of the methods discussed in Chap. 5.

Summary: concepts, tools, and procedures to know

Concepts and Tools

- **Particle positions and velocities of electron (hole) wavepackets in crystals under slowly varying external forces can be identified by semiclassical approximations to the quantum-mechanical quantities.** In particular, by the average $\bar{\mathbf{R}}$ over the Bloch functions and by the wavepacket group velocity $\mathbf{v}_{n,g}$ determined by the gradient of the energy band expression. The dynamical behaviors $\bar{\mathbf{R}}(t)$ and $\mathbf{v}_{n,g}(t)$ are determined by Newtons-like equations where the slowly-varying external fields enter as external forces in classical manner and the rapidly-varying periodic-potential field enters in quantum manner through the energy band determining $\mathbf{v}_{n,g}$.
- Under the action of a constant external force, a wavepacket in a periodic potential executes periodic oscillations within the BZ that are both in momentum and in real space. These are referred as Bloch oscillations. Their period T_B is determined by the strength of the external force and by the lattice spacing a .
- **Filled electronic bands do not contribute to electric currents: metals are distinguished from insulators, according to whether electronic bands are completely or partially filled.**
- For almost filled bands the concept of hole is conveniently introduced, as an effective particle with wavenumber, mass, energy, and charge characterized by opposite sign than that of the electrons almost filling in the band. All the previous concepts apply to holes too.
- **In the presence of a constant magnetic field, the projection $\mathbf{r}_\perp(t)$ on a plane perpendicular to \mathbf{B} of an electron trajectory in real space and their trajectory $\mathbf{k}(t)$ in \mathbf{k} -space are reciprocally connected.**
- Cyclotron motion involves two out of the three components of \mathbf{k} , thus cyclotron and band effective masses are in general different in the vicinity of critical points.
- **Occurrence of Ohm-like electric conduction, thermal conduction and related effects under the action of external fields, requires mechanisms that are able to break the crystal translational symmetry, such as lattice vibrational excitations, impurities and other defects. Scattering against fixed lattice ions does not lead to Ohm's conduction.**

- In fact, the action of external fields and of scattering mechanisms affects the deviation $\delta\mathbf{k}$ of the crystal momentum from its initial value: this is a consequence of Pauli principle.
- Scattering can be described in terms of the average time between each two scattering events, its inverse also representing the probability per unit time that a charge carrier be scattered.
- **The characteristic behaviors of materials can be probed by observing the response of the system to external disturbances:** these are conveniently chosen of the type, frequency and wavelength suited to excite a time and/or space fluctuation about the average value of that given property. The way the system responds, reveals information on interactions, statistics, and dimensionality. The temperature dependence of response coefficients is determined by the microscopic interaction driving the scattering event, as well as by statistics and dimensionality.
- When slowly-varying external disturbances act on the material, conditions of local thermodynamical equilibrium can be considered in the sample, that are set in by the high collision rate on the scale of external disturbance variations. This is the collisional or hydrodynamic regime.
- Electric conductivity measures the material response to an electric field, leading to the manifestation of an electric current. Resistivity, that is inverse conductivity, in metals depends on temperature with different power-laws: \simeq constant for $T \rightarrow 0$, $\propto T$ for $T > \Theta_D$, and $\propto T^5$ for $T \ll \Theta_D$. When impurities are present, resistivity emerges as if two independent scattering mechanisms occur, one due to impurities and one to a temperature-dependent mechanism.
- Thermal conductivity and thermoelectric coefficient measure the material response to a temperature gradient, leading to the manifestation of an energy current and of a thermoelectric field at open circuit, respectively. The response can be due to particles and excitations, such as phonons. The temperature dependence of $\chi_{c,t}$ follows the T -dependence of τ and c_V .
- **The ratio of electric-to-thermal conductivity does not depend on the scattering mechanism.**
- In metals, electron contribution to thermal conductivity is larger than the phonon one. In insulators only the phonon contribution is relevant. In semiconductors both exist, the predominance of either one of the two depending on the doping level.
- The current response $\mathbf{j} = \sigma_H \cdot \mathbf{E}$ to a combined electric and magnetic orthogonal field is represented by a tensorial-like magnetoresistance σ_H , so that electric currents set in a given direction with intensity that is determined by the electric field in another direction.
- **In materials, thermodynamic and transport properties of fermions show up as if they were determined by only a reduced number of electron states $\sim g(\epsilon_f)k_bT$ in the vicinity of the Fermi energy and with excitation energy k_bT .**

This is the essential concept brought in by the Sommerfeld theory of metals, which fixes many flaws of Drude model.

- **Extreme conditions of quantum degeneracy, such as can be met with reduced dimensions, low temperatures and strong external magnetic fields, favor the realization of new quantum states due to the occurrence of Integer and Fractional Quantum Hall Effect.**
- In Quantum Hall Effects, very peculiar quantum transport behavior emerges, such as quantized resistance and infinite conductivity, and can be used to perform precision measurements of fundamental constants.
- **In semiconductors, doping is a quite versatile tool to tune the thermodynamic and transport properties of the material.** This concept can be extraordinarily useful for applications. In extrinsic semiconductors, the mass-action law holds $n_c p_v \leq n_i^2$ whatever the position and the number of doping-induced impurities states might be. Electric conductivity in doped semiconductors is finite at finite T , its size depending on the number impurities. At higher T the conductivity drops to a smaller value that is independent of the number of impurities. At zero temperature conductivity is zero.
- **In superconductors, coherence of the pair-correlated ground state leads to sudden vanishing of the resistivity below T_c : the carriers behave as if they were one single particle not having though an available state where to scatter into.**
- In a superconductor-superconductor junction, the difference in the superconductor global phases across the junctions leads to a tunneling supercurrent of pairs even in the absence of applied voltage. This is the Josephson effect.
- In a SQUID, maximum currents can be obtained after appropriate tuning of the magnetic flux trapped by the device or else the external magnetic field.
- **At a phenomenological level, the scattering time in crystals can be written as $\tau \propto 1/n_{scatt}$, with n_{scatt} the effective number of scatterers participating to the process, either phonons or impurities. Thus, $\tau(T)$ is driven by $n_{scatt}(T)$ dependence.**
- Completely filled bands are distinguished because no electron-driven scattering, either interband or intraband, is relevant at all.
- Electron-electron scattering events are rare processes in both bulk metals and semiconductors. This may not be the case in extreme quantum conditions of e.g. reduced dimensionality or for very strong coupling interactions.
- **In $p-n$ junctions, exponentially varying currents set in after variations an external applied voltage V : the current is exponentially saturated to a vanishing value when $V < 0$ and enhanced to significant values even applying tiny positive voltage differences. This current-voltage characteristics makes the $p-n$ junction working as a current rectifier with reduced Joule dissipation because the device is small and it requires small applied voltages.**

- Thermodynamical equilibrium conditions in a $p-n$ junction manifest with the appearance of a charge depletion and of non-uniform electric potential profiles across the junction, ensuring that the chemical potential be the same within the whole device.
- Einstein relations state that a measurement of the mobility μ , as it can be easily performed in a transport-like experiment under the drift of an external electric field, provides information on the diffusion constants D that are in turn influenced by internal interactions. In fact, the Einstein relations represent a specific manifestation of a more general statement, that is that measurements of observables driven by an external field infer information on the internal system structure.
- Non-ohmic behavior characterize transistor transport behavior. The source-to-drain conductance cannot exceed $G = I/V_D = e^2/(2\pi\hbar)$ and is dictated solely by fundamental quantum constants, not by sample geometry.

Procedures

- Determination of carrier dynamical behavior within a semiclassical approach for the external electric and magnetic fields.
- Determination of the cyclotron frequency of closed orbits and of cyclotron effective mass.
- Determination of thermodynamic quantities in the Fermi-Dirac statistics and related transport properties within Sommerfeld model.
- Determination of the chemical potential in semiconductors.
- Solution of the donor or acceptor problem.

4.10 Appendix. Sommerfeld integrals

Consider the integral

$$I = \int_{-\infty}^{\infty} H(\varepsilon) f(\varepsilon) d\varepsilon, \quad (4.181)$$

with $f(\varepsilon)$ the Fermi distribution function, $H(\varepsilon)$ a function admitting derivative at all orders, and such that $\lim_{\varepsilon \rightarrow -\infty} H(\varepsilon) = 0$. Assuming that $K(\varepsilon) = \int_{-\infty}^{\varepsilon} H(\varepsilon') d\varepsilon'$ be defined and that $\lim_{\varepsilon \rightarrow \infty} H(\varepsilon) f(\varepsilon) = 0$, one finds

$$\begin{aligned} \int_{-\infty}^{\infty} H(\varepsilon) f(\varepsilon) d\varepsilon &= \int_{-\infty}^{\infty} K(\varepsilon) \left(-\frac{\partial f(\varepsilon)}{\partial \varepsilon} \right) d\varepsilon \\ K(\varepsilon) &= K(\mu) + \sum_{n=1}^{\infty} \frac{(\varepsilon - \mu)^n}{n!} \left[\frac{\partial K(\varepsilon)}{\partial \varepsilon} \right]_{\varepsilon=\mu}. \end{aligned} \quad (4.182)$$

Since $f'(\varepsilon)$ is an even function of $\varepsilon - \mu$, only those terms in the expansion of $K(\varepsilon)$ that are even in n , do contribute to the integral. One thus obtains:

$$I = \sum_{n=0}^{\infty} \frac{d^{2n}K(\varepsilon)}{d\varepsilon^{2n}} \Big|_{\varepsilon=\mu} (kT)^{2n} a_n = K(\mu) + \sum_{n=1}^{\infty} \frac{d^{2n-1}H(\varepsilon)}{d\varepsilon^{2n-1}} \Big|_{\varepsilon=\mu} (kT)^{2n} a_n, \quad (4.183)$$

where

$$a_n = \frac{1}{(2n)!} \int_{-\infty}^{\infty} x^{2n} \frac{e^x}{(e^x + 1)^2} dx.$$

4.11 Appendix. Landau levels

Consider one single electron in a magnetic field \mathbf{B} . In the Landau gauge $\mathbf{A} = B(-y, 0, 0)$ and $\nabla \cdot \mathbf{A} = 0$ for the vector potential \mathbf{A} , the Hamiltonian

$$H = \frac{1}{2m} (\mathbf{p} - \frac{e}{c} \mathbf{A})^2 = \frac{1}{2m} [p_y^2 + p_z^2 + (p_x + \frac{e}{c} By)^2]. \quad (4.184)$$

The operators p_x and p_z commute with the Hamiltonian. Therefore, the eigenfunction $\psi(r)$ of H can be cast in the form

$$\psi(\mathbf{r}) = e^{ik_x x} e^{ik_z z} F(y). \quad (4.185)$$

Simple substitution of this expression into $H\psi = \varepsilon\psi$ leads to the result that $F(y)$ is an eigenfunction of the equation for the harmonic oscillator:

$$-\frac{\hbar^2}{2m} \frac{d^2 F(y)}{dy^2} + \frac{1}{2} m \omega_c^2 (y - y_0)^2 F(y) = \varepsilon' F(y). \quad (4.186)$$

Here, $\varepsilon' = \varepsilon - \hbar^2 k_z^2 / (2m)$ is defined along with the cyclotron frequency $\omega_c = eB/(mc)$, and $y_0 = -c\hbar k_x / (eB)$. The spectrum is given by

$$\varepsilon = \frac{\hbar^2}{2m} k_z^2 + (n + \frac{1}{2}) \hbar \omega_c. \quad (4.187)$$

This is the sum of a single-particle parabolic term running on top of a discrete term, as shown in Fig. 4.20. The discrete structure is dictated by orbits quantization. The spectrum does not depend on y_0 , and therefore each energy level is degenerate, with degeneracy equal to the number of possible values of y_0 . Labeled L the system size along y , one obtains $0 < |y_0| < L$ and therefore $0 < k_x < eBL/(c\hbar)$. Since the number of states in the interval dk_x is given by $L(2\pi)^{-1} dk_x$, the degeneracy of an energy level with given k_z and n is $L(2\pi)^{-1} dk_x^{max} / L^2$ or:

$$d = \frac{L}{2\pi} \frac{eBL}{c\hbar} \frac{1}{L^2} = \frac{eB}{2\pi c\hbar}. \quad (4.188)$$

Considering that the number of discrete states in a unit of energy is $1/(\hbar\omega_c)$, one finds that the density of states for given k_z is $mL^2/(2\hbar^2\pi)$. Notice that the latter expression coincides with that of a system of free electrons in two dimensions. The density of states for given n is

$$G(\varepsilon) = d \frac{L}{2\pi} \frac{dk_z}{d\varepsilon} = \frac{eB\sqrt{2m}}{8\pi^2\hbar^3 c \sqrt{\varepsilon - \varepsilon_n}}, \quad (4.189)$$

with $\varepsilon_n = (n + 1/2)\hbar\omega_c$.

To summarize, the spectrum (4.187) is characterized by the presence of two terms on top of each other: the continuous term $\hbar^2 k_z^2 / (2m)$ representing the kinetic energy of a free particle along the direction of the magnetic field, and the discrete term $(n + 1/2)\hbar\omega_c$ representing the energy of a 1D harmonic oscillator whose oscillation center lies in the x-y plane. The density of states behaves as if the states be re-distributed within the otherwise unchanged number of particles. The redistribution manifests as illustrated in Fig. 4.21, with peaked structures located at the discrete values of the energy spectrum.

In the case in which the electron is bound to stay onto the plane orthogonal to the magnetic field, the resulting spectrum is that of a harmonic oscillator.

4.12 Appendix. Fermi-Dirac integrals

The general definition of Fermi-Dirac integrals is

$$\tilde{F}_j(\eta_f) = \frac{1}{\Gamma(j+1)} \int_0^\infty dx \frac{x^j}{1 + e^{x-\eta_f}}, \quad (4.190)$$

with $\Gamma(x)$ the Gamma function. For $j = 1/2$ and $\eta_f = (\mu - \varepsilon_c)/(k_b T)$ one finds that the number of electrons per unit volume in the conduction band n_c , is given by

$$\begin{aligned} n_c &= N_{3d} \frac{2}{\sqrt{\pi}} F_{\frac{1}{2}}(\eta_f) = N_{3d} \tilde{F}_{\frac{1}{2}}(\eta_f) \\ N_{3d} &= 2 \left(\frac{2\pi m_c^* k_b T}{\hbar^2} \right)^{\frac{3}{2}}, \end{aligned} \quad (4.191)$$

as it is seen from (4.110). In the case of one- and two-dimensional systems one finds:

$$\begin{aligned} n_{2d} &= N_{2d} \tilde{F}_0(\eta_f) \\ N_{2d} &= \frac{m_c^* k_b T}{\pi \hbar^2} \\ n_{1d} &= N_{1d} \tilde{F}_{-\frac{1}{2}}(\eta_f) \\ N_{1d} &= \frac{1}{\hbar} \sqrt{\frac{2m_c^* k_b T}{\pi}}. \end{aligned} \quad (4.192)$$

The properties of Fermi-Dirac integrals are summarized as follows:

Properties

P1.

$$\begin{aligned}\tilde{F}_0(\eta_f) &= \log(1 + e^{\eta_f}) \\ \frac{d\tilde{F}_j(\eta_f)}{d\eta_f} &= \tilde{F}_{j-1}(\eta_f).\end{aligned}\quad (4.193)$$

P2. In the non-degenerate case with $\eta_f \ll 0$, the asymptotic behavior

$$\tilde{F}_j(\eta_f) = e^{\eta_f} \quad (4.194)$$

sets in.

P3. In the degenerate case with $\eta_f \gg 0$, one has

$$\begin{aligned}\tilde{F}_j(\eta_f) &= 2\eta_f^{j+1} \sum_{n=0}^{\infty} \frac{t_{2n}}{\Gamma(j+2-2n)\eta_f^{2n}} \\ &\quad + \cos(\pi j) \sum_{n=1}^{\infty} \frac{(-1)^{n-1} e^{-n\eta_f}}{n^{j+1}}\end{aligned}\quad (4.195)$$

with

$$\begin{aligned}t_0 &= \frac{1}{2} \\ t_n &= \sum_{p=1}^{\infty} \frac{(-1)^{p-1}}{p^n} = (1 - 2^{1-n}) \zeta(n),\end{aligned}\quad (4.196)$$

and $\zeta(n)$ the Riemann zeta function.

P4. Selected case functions are:

$$\begin{aligned}\tilde{F}_{-\frac{1}{2}}(\eta_f) &= \frac{2\eta_f^{\frac{1}{2}}}{\sqrt{\pi}} & \tilde{F}_{\frac{1}{2}}(\eta_f) &= \frac{4\eta_f^{\frac{3}{2}}}{3\sqrt{\pi}} & \tilde{F}_1(\eta_f) &= \frac{1}{2}\eta_f^2 \\ \tilde{F}_{\frac{3}{2}}(\eta_f) &= \frac{8\eta_f^{\frac{5}{2}}}{15\sqrt{\pi}} & \tilde{F}_2(\eta_f) &= \frac{1}{6}\eta_f^3.\end{aligned}\quad (4.197)$$

4.13 Appendix. Fermi-Dirac statistics for impurity states in semiconductors

The self-consistent calculation of the electronic levels leads to the conclusion that each level can be occupied by no more than two electrons. In particular, if only one state is taken into account, the latter can be occupied in either one of the following manners: (a) simply occupied with $N = 0$ and $E = 0$; (b) occupied in two different manners by only one electron, thus with $N = 1$ and $E = E_d = \epsilon_c + \epsilon_{d,1}$; (c) occupied by two electrons with opposite spins, therefore with $N = 2$ and $E = 2E_d$.

When the system can exchange energy and particles with its environment, one has that the average number of particles at thermodynamical equilibrium is given by

$$\langle n \rangle = \frac{\sum_{m,N} N e^{-\beta(E_{m,N} - \mu N)}}{\sum_{m,N} e^{-\beta(E_{m,N} - \mu N)}}, \quad (4.198)$$

where $E_{m,N}$ is the energy of the m -th system state with N particles, μ is the chemical potential and $\beta = 1/(k_b T)$. Therefore, it turns out that

$$\begin{aligned} \langle n \rangle &= \frac{e^{-\beta(E_d - \mu)} + e^{-\beta(E_d - \mu)} + 2e^{-\beta(2E_d - 2\mu)}}{1 + e^{-\beta(E_d - \mu)} + e^{-\beta(E_d - \mu)} + e^{-\beta(2E_d - 2\mu)}} \\ &= 2e^{-\beta(E_d - \mu)} \frac{1 + e^{-\beta(E_d - \mu)}}{(1 + e^{-\beta(E_d - \mu)})^2} = 2 \frac{1}{1 + e^{\beta(E_d - \mu)}}. \end{aligned} \quad (4.199)$$

The Fermi distribution function is then obtained, with an additional factor of two that accounts for spin degeneracy.

If the level represents a donor state, the possible configurations are (a) and (b), whereas (c) is not since electrostatic repulsion prevents two electrons from staying onto the same level. In this case:

$$\begin{aligned} \langle n \rangle &= \frac{e^{-\beta(E_d - \mu)} + e^{-\beta(E_d - \mu)}}{1 + e^{-\beta(E_d - \mu)} + e^{-\beta(E_d - \mu)}} \\ &= \frac{1}{\frac{1}{2}e^{\beta(E_d - \mu)} + 1} = P_e(E_d). \end{aligned} \quad (4.200)$$

For symmetric reasons, if the level represents an acceptor state, the possible configurations are (b) and (c), whereas (a) is not since it would imply the presence of two holes onto the level. In this case:

$$\begin{aligned} \langle n \rangle &= \frac{e^{-\beta(E_a - \mu)} + e^{-\beta(E_a - \mu)} + 2e^{-\beta(2E_a - 2\mu)}}{e^{-\beta(E_a - \mu)} + e^{-\beta(E_a - \mu)} + e^{-\beta(2E_a - 2\mu)}} \\ &= 2 - \frac{1}{1 + \frac{1}{2}e^{\beta(\mu - E_a)}}. \end{aligned} \quad (4.201)$$

The occupation probability for a hole of the acceptor level clearly results to be

$$P_h(E_a) = \frac{1}{1 + \frac{1}{2}e^{\beta(\mu - E_a)}}. \quad (4.202)$$

4.14 Appendix. Hall coefficient in semiconductors

Starting from (4.142) and (4.143), the total conductivity tensor $\hat{\sigma} = \hat{\sigma}_1 + \hat{\sigma}_2$ can be easily calculated. The corresponding components are:

$$\begin{aligned} \sigma_{xx} &= \sigma_{yy} = \frac{\sigma_1}{P_1} + \frac{\sigma_2}{P_2} \\ \sigma_{xy} &= -\sigma_{yx} = -\frac{\sigma_1 \omega_{1c} \tau_1}{P_1} + \frac{\sigma_2 \omega_{2c} \tau_2}{P_2}, \end{aligned} \quad (4.203)$$

with $P_1 = 1 + (\omega_{1c} \tau_1)^2$ and $P_2 = 1 + (\omega_{2c} \tau_2)^2$. Inserting the explicit expressions for σ_1 and σ_2 , the relation between current and electric field can be cast in the form:

$$\begin{aligned} j_x &= \sigma_{xx} E_x + \sigma_{xy} E_y \\ j_y &= -\sigma_{xy} E_x + \sigma_{xx} E_y, \end{aligned} \quad (4.204)$$

with

$$\begin{aligned} \sigma_{xx} &= e^2 \left(\frac{n_1 \mu_1}{P_1} + \frac{n_2 \mu_2}{P_2} \right) \\ \sigma_{xy} &= e^2 \left(-\frac{n_1 \mu_1 \omega_{1c} \tau_1}{P_1} + \frac{n_2 \mu_2 \omega_{2c} \tau_2}{P_2} \right), \end{aligned} \quad (4.205)$$

and $\mu_1 = \tau_1/m_1$ and $\mu_2 = \tau_2/m_2$.

Under conditions of zero current $j_y = 0$, one has that $E_y = \sigma_{xy} E_x / \sigma_{xx}$ and that $j_x = (\sigma_{xx} + \sigma_{xy}^2 / \sigma_{xx}) E_x$. In this case, the Hall coefficient is defined by

$$R_H = \frac{E_y}{j_x B} = \frac{\sigma_{xy}}{(\sigma_{xx}^2 + \sigma_{xy}^2) B}. \quad (4.206)$$

Simple algebra leads to the result

$$\begin{aligned} \sigma_{xx}^2 + \sigma_{xy}^2 &= e^4 \left[\frac{(n_1 \mu_1)^2}{P_1} + \frac{(n_2 \mu_2)^2}{P_2} \right. \\ &\quad \left. + \frac{2n_1 \mu_1 n_2 \mu_2}{P_1 P_2} (1 - \omega_{1c} \tau_1 \omega_{2c} \tau_2) \right]. \end{aligned} \quad (4.207)$$

It is therefore seen that the sign of the Hall coefficient is the same as the sign of σ_{xy} . In particular, $\sigma_{xy} = 0$ implies $E_y = 0$. End-of-chapter problems are devoted to determine the dependence of resistivity on the magnetic field.

4.15 Appendix. Currents from minority carriers in $p-n$ junctions

The currents of electrons and holes depend on the electric field. Since the charge distribution is non homogeneous, a current is expected to set in, flowing from regions with larger to regions with lower densities. This is a diffusive-like contribution to the current, that can be represented by the expressions $-D_n dn/dx$ and $-D_p dp/dx$ for electrons and holes respectively: the current is proportional to the density gradient through the transport coefficients D_n and D_p . These are the so-called Einstein diffusion coefficients. Notice that the current has opposite sign with respect to the density gradient dn/dx , since it tends to restore a uniform density $n(x)$. Overall, the currents can be cast in the form:

$$\begin{aligned} J_e(x) &= -\mu_n n E - D_n \frac{dn}{dx} \\ J_h(x) &= \mu_p p E - D_p \frac{dp}{dx}, \end{aligned} \quad (4.208)$$

where μ_n and μ_p are defined as in (4.139). Both (4.208) state that the total current of electrons and holes is the sum of two terms, one in the form of a Ohm's process driven by the electric field, and one due to diffusion processes originated by inhomogeneities in the local density.

Two opposite limiting behaviors can be discussed. In the region with $|D_n dn/dx| \ll |\mu_n n E|$, the diffusive term is not significant and therefore the relationship between current and electric field is essentially the Ohm's law. On the other hand, in the region with $|D_n dn/dx| \gg |\mu_n n E|$, the current is mainly diffusive-like.

It is immediately verified that when n and p are given by (4.151), that is under conditions of thermodynamical equilibrium, one has:

$$\begin{aligned} \frac{dn}{dx} &= -\frac{neE}{k_b T} \\ \frac{dp}{dx} &= \frac{peE}{k_b T}. \end{aligned} \quad (4.209)$$

As a consequence, $J_e(x) = J_h(x) = 0$ when

$$\begin{aligned} -\mu_n + \frac{eD_n}{k_b T} &= 0 \\ \mu_p - \frac{eD_p}{k_b T} &= 0. \end{aligned} \quad (4.210)$$

Concept

These are known as Einstein relations. They state that a measurement of the mobility μ , as it can be easily performed in a transport-like experiment under the drift of an external electric field, provides information on the diffusion constants D that are in turn influenced by internal interactions. In fact, the Einstein relations represent a specific manifestation of a more general statement, that is that measurements of observables driven by an external field infer information on the internal system structure.

Viceversa, inserting in (4.208) the relations $J_e = 0$ and $J_h = 0$, and the expressions (4.210) for D_n e D_p , one has:

$$\begin{aligned}\frac{dn}{dx} &= -\frac{\mu_n n E}{D_n} = -\frac{en}{k_b T} \frac{d\phi(x)}{dx} \\ \frac{dp}{dx} &= -\frac{\mu_p p E}{D_p} = -\frac{ep}{k_b T} \frac{d\phi(x)}{dx}.\end{aligned}\quad (4.211)$$

After integrating over x , this yields

$$\begin{aligned}n(x) &= n_0 e^{\frac{e\phi(x)}{k_b T}} \\ p(x) &= p_0 e^{-\frac{e\phi(x)}{k_b T}}.\end{aligned}\quad (4.212)$$

These expressions coincide with the relations (4.151) if

$$\begin{aligned}n_0 &= N_c e^{-\frac{\varepsilon_c - \mu}{k_b T}} \\ p_0 &= P_v e^{\frac{\mu - \varepsilon_v}{k_b T}}.\end{aligned}\quad (4.213)$$

This means that (4.208) embodies as special cases the equilibrium distributions of electrons and holes in the conduction and valence bands. The number of unknowns in (4.208), that are μ , $n(x)$, $p(x)$ and $\phi(x)$, exceeds the number of equations, and therefore two additional equations are needed: these are the relations expressing the conservation of the number of electrons and of the number of holes. In conventional one-dimensional systems this relation takes the form

$$\begin{aligned}\frac{\partial J_e}{\partial x} &= -\frac{\partial n}{\partial t} \\ \frac{\partial J_h}{\partial x} &= -\frac{\partial p}{\partial t},\end{aligned}\quad (4.214)$$

that is of a continuity equation. Here, the following possibilities are to be accounted for an electron in the conduction band: (a) it can be either captured or generated by an impurity with either non-occupied or occupied state, respectively; (b) it can recombine with a hole; (c) it can be generated by creation of an electron-hole pair.

Similar possibilities are to be accounted for a hole in the valence band. Equations (4.214) do not embody the latter processes, so that new terms are to be introduced. The whole set of new terms are represented by $(dn/dt)_{g \rightarrow r}$ and $(dp/dt)_{g \rightarrow r}$, so that the conservation for the number of electrons and holes can now be cast in the form:

$$\begin{aligned}\frac{\partial n}{\partial t} &= -\frac{\partial J_e}{\partial x} + \left(\frac{dn}{dt}\right)_{g \rightarrow r} \\ \frac{\partial p}{\partial t} &= -\frac{\partial J_h}{\partial x} + \left(\frac{dp}{dt}\right)_{g \rightarrow r}.\end{aligned}\quad (4.215)$$

The detailed description of all recombination and generation processes is complex: a more simple treatment can be developed at a phenomenological level by introducing average lifetimes $\bar{\tau}_n$ and $\bar{\tau}_p$ for electrons and holes. One can then write:

$$\begin{aligned}\left(\frac{dn}{dt}\right)_{g \rightarrow r} &= -\frac{n - n_c}{\bar{\tau}_n} \\ \left(\frac{dp}{dt}\right)_{g \rightarrow r} &= -\frac{p - p_v}{\bar{\tau}_p}.\end{aligned}\quad (4.216)$$

The two (4.216) describe relevant processes. For example, if at a given time $n > n_c$, with n_c the number of electrons in conduction band at thermodynamical equilibrium, one has that $(dn/dt)_{g \rightarrow r}$ is negative: as a result, the electronic distribution tends towards equilibrium, while decreasing the number of electrons per unit time by the amount $(n - n_c)/\bar{\tau}_n$. Similarly, if $n < n_c$ instead, $(dn/dt)_{g \rightarrow r}$ is positive and the distribution tends towards equilibrium while increasing the number of electrons per unit time by the amount $(n - n_c)/\bar{\tau}_n$. The two introduced times $\bar{\tau}_n$ and $\bar{\tau}_p$ govern the generation and recombination processes of electrons and holes. They are to be conceptually distinguished with respect to the collisional times τ_n and τ_p : in fact, $\bar{\tau}_n$ and $\bar{\tau}_p$ are as long as $10^{-3} \div 10^{-8}$ s, significantly longer than the $\approx 10^{-13}$ s values typical of τ_n and τ_p .

Selected consequences of (4.215) can be discussed under stationary conditions, that is $\partial n/\partial t = \partial p/\partial t = 0$. From (4.215) and (4.216), one obtains:

$$\begin{aligned}\frac{\partial J_e}{\partial x} + \frac{n - n_c}{\bar{\tau}_n} &= 0 \\ \frac{\partial J_h}{\partial x} + \frac{p - p_v}{\bar{\tau}_p} &= 0.\end{aligned}\quad (4.217)$$

A special case exists for which the relations (4.217) can be solved. This is the limit in which the diffusive terms in (4.208) prevail. In this limit, one obtains

$$\begin{aligned}D_n \frac{\partial^2 n}{\partial x^2} &= \frac{n - n_c}{\bar{\tau}_n} \\ D_p \frac{\partial^2 p}{\partial x^2} &= \frac{p - p_v}{\bar{\tau}_p}.\end{aligned}\quad (4.218)$$

The corresponding solutions are

$$\begin{aligned} n(x) &= n_c + n_0 e^{-\frac{x}{L_n}} \\ p(x) &= p_v + p_0 e^{-\frac{x}{L_p}}. \end{aligned} \quad (4.219)$$

Here, n_0 and p_0 are the initial values of the density of electrons and holes and

Definition

$L_n = (D_n \bar{\tau}_n)^{\frac{1}{2}}$ and $L_p = (D_p \bar{\tau}_p)^{\frac{1}{2}}$ are the so-called diffusion lengths.

For example, combining the Einstein relations with the Drude expression for the mobility $\mu_n = e\tau_n/m_n$, the equipartition theorem $3k_bT = m_n v_{n,th}^2$, and the definition of mean-free path $l_n = \bar{\tau}_n v_{n,th}$, one obtains:

$$L_n = \sqrt{\frac{k_b T \tau_n \bar{\tau}_n}{m}} = \sqrt{\frac{v^2 \tau_n \bar{\tau}_n}{3}} l_n \sqrt{\frac{\bar{\tau}_n}{3 \tau_n}}. \quad (4.220)$$

Notice that the ratio $\bar{\tau}_n/\tau_n \equiv N_{u,n}$ measures the number of collisions during the time $\bar{\tau}_n$. Thus, (4.220) states that $L_n = l_n \sqrt{N_{u,n}}$, that is the diffusion length coincides with the distance traveled after $N_{u,n}$ steps with length l_n in its random walk. A similar reasoning can be repeated for holes, leading to $L_p = l_p \sqrt{N_{u,p}}$.

The currents of minority carriers are to be calculated from (4.164), using the following considerations: given the n -zone for $x > d_n$ and starting from point $x_0 \geq d_n$, one assumes that the minority carriers reach the asymptotic limit for values $x > d_n$. Thus, the distribution function for the minority carriers results to be:

$$p(x) = p(\infty) + (p(x_0) - p(\infty)) e^{-\frac{x-x_0}{L_p}}, \quad (4.221)$$

with $p(\infty) = n_i^2/N_c$ and $x > x_0$. A similar reasoning can be followed in the p -region for $x < -d_p$, leading to

$$n(x) = n(-\infty) + (n(x'_0) - n(-\infty)) e^{\frac{x-x'_0}{L_n}}, \quad (4.222)$$

where $n(-\infty) = n_i^2/N_p$, $x'_0 < -d_p$, and $x < x'_0$.

The assumptions $p(x_0) \neq p(\infty)$ and $n(x'_0) \neq n(-\infty)$ can be understood as follows. In the case with $J_e = J_h = 0$ in (4.217), the diffusive terms disappear. In all the other cases, (4.221) and (4.222) state that, taking $x_0 = d_n$ and $x'_0 = -d_p$, L_n and L_p measure the size of the region to the right of d_n and to the left of $-d_p$, where the distribution of minority carriers reaches the asymptotic values.

Based on the above results, J_e^{gen} and J_h^{gen} can be evaluated: in the n -region, $p_0(\infty)/\bar{\tau}_p$ holes are thermally generated per unit of time and of volume. All the holes generated along one diffusion length L_p are very likely to be channeled in the depletion region. Therefore, the currents of minority carriers, either electrons or holes, turn out to be:

$$\begin{aligned}
 J_h^{gen} &= \frac{n_i^2}{N_d} \frac{L_p}{\bar{\tau}_p} \\
 J_e^{gen} &= \frac{n_i^2}{N_a} \frac{L_n}{\bar{\tau}_n}.
 \end{aligned} \tag{4.223}$$

In conclusion, one finds that

$$J_h^{gen} + J_e^{gen} = \frac{n_i^2}{N_d} \frac{L_p}{\bar{\tau}_n} + \frac{n_i^2}{N_a} \frac{L_n}{\bar{\tau}_n}. \tag{4.224}$$

Problems with solutions

4.1. Given the equations of motion (4.36), find the cyclotron mass (4.37).

Solution. In the reference system with axes along the directions of the principal effective masses, the magnetic field has components $(B \sin \theta \cos \varphi, B \sin \theta \sin \varphi, B \cos \theta)$. Defining $\mathbf{v} = u e^{-i\omega t}$, from (4.36) one obtains the system of linear homogeneous equations

$$\begin{aligned}
 -i\omega u_x m_1 + \frac{eB}{c} (u_y \cos \theta - u_z \sin \theta \sin \varphi) &= 0 \\
 -\frac{eB}{c} u_x \cos \theta - i\omega u_y m_2 + \frac{eB}{c} u_z \sin \theta \cos \varphi &= 0 \\
 \frac{eB}{c} (u_x \sin \theta \sin \varphi - u_y \sin \theta \cos \varphi) - i\omega u_z m_3 &= 0.
 \end{aligned} \tag{4.225}$$

Given $\omega = eB/(cm^*)$, the condition that the determinant of the coefficients vanishes provides the relation (4.37).

4.2. Deduce (4.64) from (4.63).

Solution. The functions $\varepsilon(x \pm \ell_x)$ can be expanded in Taylor series to first order in $\pm \ell_x$, yielding $\varepsilon(x \pm \ell_x) = \varepsilon(x) \pm \ell_x d\varepsilon(x)/(dx)$. Substituting into (4.63), the terms having $\varepsilon(x)$ as a factor yield the electric current, assumed to vanish, except for the factor e . One thus obtains the energy current:

$$\begin{aligned}
 j_e &\simeq -\ell_x \frac{d\varepsilon(x)}{dx} \frac{n}{2} [v_d + v_x(x - \ell_x)] + \ell_x \frac{d\varepsilon(x)}{dx} \frac{n}{2} [v_d - v_x(x + \ell_x)] \\
 &\simeq -\ell_x \frac{d\varepsilon(x)}{dx} n v_x(x).
 \end{aligned} \tag{4.226}$$

Besides, the electric current results to be

$$\begin{aligned}
 0 &\simeq -\frac{n}{2} e [v_d + v_x(x) - \ell_x \frac{dv_x(x)}{dx}] - \frac{n}{2} e [v_d - v_x(x) - \ell_x \frac{dv_x(x)}{dx}] \\
 &= -nev_d + ne\ell_x \frac{dv_x(x)}{dx}.
 \end{aligned} \tag{4.227}$$

Relations (4.64) are deduced from these equations.

4.3. Prove that the energy per unit volume of the electron gas is given by (4.87).

Solution. The density of states is $g(\epsilon) = C\sqrt{\epsilon}$ with $C = (2m)^{\frac{3}{2}}/(2\pi^2\hbar^3)$. After integration by part, the energy $U(T)$ per unit volume is

$$U(T) = C \int_0^\infty \epsilon^{\frac{3}{2}} f(\epsilon) d\epsilon = C \left[\frac{2}{5} \epsilon^{\frac{5}{2}} f(\epsilon) \Big|_0^\infty - \frac{2}{5} \int_0^\infty \epsilon^{\frac{5}{2}} f'(\epsilon) d\epsilon \right].$$

Introducing the variable $x = (\epsilon - \mu)/(k_b T)$, given $\mu/(k_b T) \gg 1$ and expanding $(\mu + k_b T x)^{\frac{5}{2}}$ to the second order in $k_b T/\mu$, one obtains

$$\begin{aligned} U(T) &= -\frac{2}{5} C \int_{-\frac{\mu}{k_b T}}^\infty (\mu + k_b T x)^{\frac{5}{2}} \frac{df(x)}{dx} dx \\ &\simeq -\frac{2}{5} C \int_{-\infty}^\infty dx \mu^{\frac{5}{2}} \left[1 + \frac{5}{2} \frac{k_b T}{\mu} x + \frac{15}{8} \left(\frac{k_b T}{\mu} \right)^2 x^2 \right] \frac{df(x)}{dx}. \end{aligned} \quad (4.228)$$

Since $df(x)/dx$ is an even function, one has

$$U(T) = \frac{2}{5} C \left[\mu^{\frac{5}{2}} + \frac{15}{8} \frac{\pi^2}{3} \mu^{\frac{1}{2}} (k_b T)^2 \right],$$

with

$$\frac{\pi^2}{3} = - \int_{-\infty}^\infty dx x^2 \frac{df(x)}{dx}.$$

Since μ is given by (4.86), after direct substitution one obtains

$$\begin{aligned} U(T) &= \frac{2}{5} C \left[\epsilon_f^{\frac{5}{2}} \left(1 - \frac{\pi^2}{12} \left(\frac{k_b T}{\epsilon_f} \right)^2 \right)^{\frac{5}{2}} + \frac{15}{8} \frac{\pi^2}{3} \epsilon_f^{\frac{1}{2}} (k_b T)^2 \right] \\ &\simeq \frac{3}{5} n \epsilon_f + C \frac{\pi^2}{6} \sqrt{\epsilon_f} (k_b T)^2 \\ &= \frac{3}{5} n \epsilon_f + \frac{\pi^2}{6} g(\epsilon_f) (k_b T)^2. \end{aligned} \quad (4.229)$$

4.4. Determine the behavior of the chemical potential for an electron gas in one, two, and three dimensions.

Solution. The density of states for free particles for 1D, 2D, and 3D cases is, respectively:

$$\begin{aligned} g_1(\epsilon) &= \frac{1}{\pi} \sqrt{\frac{m}{2\hbar^2}} \frac{1}{\sqrt{\epsilon}} \\ g_2(\epsilon) &= \frac{1}{2\pi} \frac{m}{\hbar^2} \end{aligned}$$

$$g_3(\epsilon) = \frac{m}{\pi^2 \hbar^2} \sqrt{\frac{2m}{\hbar^2}} \sqrt{\epsilon}. \quad (4.230)$$

The 3D case has already been worked out in the main text, obtaining:

$$\mu_3 = \epsilon_f - \frac{\pi^2}{12} \frac{(k_b T)^2}{\epsilon_f}.$$

Following the same procedure as in the 3D case, one finds:

$$\begin{aligned} \mu_2 &= \epsilon_f \\ \mu_1 &= \epsilon_f + \frac{\pi^2}{12} \frac{(k_b T)^2}{\epsilon_f}. \end{aligned} \quad (4.231)$$

4.5. Be W the height of the constant potential barrier within which electrons in a metal freely move, ϵ_f the Fermi energy, and $\phi = W - \epsilon_f$ the working function. Demonstrate that thermal effects lead to the electric current density

$$J = AT^2 e^{-\frac{\phi}{k_b T}},$$

with $A = me k_b^2 / (2\pi^2 \hbar^3)$, known in fact the Richardson-Dushman law.

Solution. The number of electrons per unit volume with momentum $\hbar k_x$ and potential W is given by

$$n = \frac{2}{(2\pi)^3} \int_{-\infty}^{\infty} dk_y \int_{-\infty}^{\infty} dk_z \frac{1}{e^{\beta \left[\frac{\hbar^2}{2m} (k_x^2 + k_y^2 + k_z^2) - \mu + \phi + \epsilon_f \right]} + 1}.$$

Since $e^{\beta(\phi + \epsilon_f)} \gg 1$ and $\mu \simeq \epsilon_f$, by integration over k_y and k_z one obtains:

$$n = \frac{1}{2\pi^2} \frac{m}{\beta \hbar^2} e^{-\beta \phi} e^{-\beta \frac{\hbar^2}{2m} k_x^2}.$$

The current density along the positive x-axis is

$$\begin{aligned} j &= \frac{1}{2\pi^2} \frac{m}{\beta \hbar^2} e^{-\beta \phi} \int_0^{\infty} dk_x \frac{\hbar k_x}{m} e^{-\beta \frac{\hbar^2}{2m} k_x^2} \\ &= \frac{1}{2\pi^2} \frac{me k_b^2}{\hbar^3} T^2 e^{-\beta \phi}. \end{aligned} \quad (4.232)$$

4.6. The electric conductivity tensor is by definition $j_i = \sum_{k=1}^3 \sigma_{ik} E_k$, where j_i and E_k are the i-th and k-th component of the current density and of the electric field, respectively. The resistivity tensor is defined by means of the inverse relation $E_i = \sum_{k=1}^3 \rho_{ik} j_k$. Both are second rank tensors. If C is the conductivity and R the resistivity tensor, demonstrate that $R = C^{-1}$.

Solution. Substitute $j_k = \sum_{p=1}^3 \sigma_{kp} E_p$ in $E_i = \sum_{k=1}^3 \rho_{ik} j_k$. One obtains

$$E_i = \sum_{k=1}^3 \rho_{ik} j_k \sum_{p=1}^3 \sigma_{kp} E_p = \sum_{p=1}^3 \left(\sum_{k=1}^3 \rho_{ik} \sigma_{kp} \right) E_p.$$

For the latter to be E_i , it is needed that $\sum_{k=1}^3 \rho_{ik} \sigma_{kp} = \delta_{ip}$, that is $R = C^{-1}$.

4.7. The thermal coefficient of linear expansion for a metallic material is $\lambda = 2 \cdot 10^{-6}/K$ constant in the whole temperature range up to 300 K. Consider the Fermi energy ε_f , temperature T_f , velocity v_f , and wavevector k_f at $T = 0$. Estimate the variation of these quantities at 300 K.

Solution. Be a the lattice parameter at $T = 0$. At room temperature one has $a' = a(1 + \lambda T)$, so that the cell volume becomes $V' = V(1 + 3\lambda T)$. Therefore, the electronic density at room temperature is

$$n' = \frac{N}{V'} = \frac{n}{1 + 3\lambda T} \simeq n(1 - 3\lambda T).$$

As a consequence, the Fermi wavenumber and velocity become

$$\begin{aligned} k'_f &= (3\pi^2 n')^{\frac{1}{3}} = k_f(1 - 3\lambda T)^{\frac{1}{3}} \simeq k_f(1 - \lambda T) \\ v'_f &= \frac{\hbar k'_f}{m} = v_f(1 - \lambda T). \end{aligned} \quad (4.233)$$

Finally, the Fermi energy is

$$\varepsilon'_f = \frac{\hbar^2}{2m} k'^2_f = \varepsilon_f(1 - \lambda T)^2 \simeq \varepsilon_f(1 - 2\lambda T).$$

Since $\lambda \sim 10^{-6}/K$, the fractional changes of n , k_f , v_f , and ε_f are $\Delta n/n = (n' - n)/n \simeq -3\lambda T \simeq -1.6 \cdot 10^{-3}$, $\Delta k_f/k_f = \Delta v_f/v_f \simeq -6 \cdot 10^{-4}$, and $\Delta \varepsilon_f/\varepsilon_f \simeq -12 \cdot 10^{-4}$, respectively.

4.8. Consider a quantum system with two degenerate energy levels, the lowest with energy 0 and degeneracy g_1 and the higher with energy ΔE and degeneracy g_2 . Calculate the specific heat C and study it as a function of temperature. Show that C vanishes at $T \rightarrow 0$ and $T \rightarrow +\infty$, getting a maximum value at intermediate temperatures. Extend the results to a three-level system with energies 0, ΔE and $2\Delta E$, and show that the specific heat vanishes at $T = 0$ and $T \rightarrow +\infty$ as well.

Solution. In the canonical ensemble, the probability for the occupation of the two energy states are, respectively:

$$\begin{aligned} p_1 &= \frac{g_1}{g_1 + g_2 e^{-\beta \Delta E}} \\ p_2 &= \frac{g_2 e^{-\beta \Delta E}}{g_1 + g_2 e^{-\beta \Delta E}} = \frac{g_2}{g_1 e^{\beta \Delta E} + g_2}. \end{aligned} \quad (4.234)$$

The number of particles in the excited state is

$$\frac{N_1}{N} = \frac{g_2}{g_1 e^{\beta \Delta E} + g_2}.$$

The energy of the system is therefore:

$$U = N_1 \Delta E = \frac{N g_2 \Delta E}{g_1 e^{\beta \Delta E} + g_2}.$$

Finally, the specific heat is

$$\begin{aligned} C &= \frac{\partial U}{\partial T} = -\frac{1}{k_b T^2} \frac{\partial U}{\partial \beta} \\ &= N k_b (\beta \Delta E)^2 \frac{g_1}{g_2} \frac{e^{\beta \Delta E}}{\left(1 + \frac{g_1}{g_2} e^{\beta \Delta E}\right)^2}. \end{aligned} \quad (4.235)$$

Setting $x = \beta \Delta E$ and $a = g_1/g_2$, one has

$$C = N k_b a \frac{x^2 e^x}{[1 + a e^x]^2}.$$

By inspection, the specific heat vanishes while $T \rightarrow 0$ and $T \rightarrow +\infty$. Since C is everywhere positive, it must also have a maximum at intermediate temperatures. Defining

$$f(x) = \frac{x^2 e^x}{[1 + a e^x]^2},$$

one has

$$f'(x) = \frac{x e^x}{[1 + a e^x]^2} \left(2 + x - \frac{2 a x e^x}{1 + a e^x} \right),$$

that vanishes when

$$2 + 2 a e^x + x - a x e^x = 0,$$

or else

$$2 + x + a(2 - x) e^x = 0.$$

The solution x_m to the above equation has to satisfy the condition $x_m > 2$. One then has

$$e^{x_m} = \frac{1}{a} \frac{2 + x_m}{x_m - 2},$$

and the specific heat becomes

$$C = \frac{N k_b}{4} \left[\left(\frac{\Delta E}{k_b T_m} \right)^2 - 4 \right].$$

In the case of three levels with degeneracy g_1 , g_2 , and g_3 , one may follow the same procedure and find:

$$U = N\Delta E \left[\frac{g_2}{g_1 e^x + g_2 + g_3 e^{-x}} + \frac{2g_3}{g_1 e^{2x} + g_2 e^x + g_3} \right].$$

The specific heat is determined from the temperature derivative as

$$C = Nk_b x^2 F(x),$$

with $F(x)$ finite at $x = 0$ and exponentially decreasing as $x \rightarrow +\infty$. The predicted result follows.

4.9. Calculate the compressibility of an electron gas at $T = 0$.

Solution. At $T = 0$ the energy is

$$E_0 = \frac{\hbar^2 V}{10m\pi^2} \left(3\pi^2 \frac{N}{V} \right)^{5/3}.$$

Using the definitions, the pressure and the compressibility are, respectively:

$$p = -\frac{\partial E_0}{\partial V} = \frac{2}{3} \frac{\hbar^2 V^{-5/3}}{10m\pi^2} (3\pi^2 N)^{5/3}$$

and

$$B = -V \left(\frac{\partial p}{\partial V} \right)_T = \frac{1}{9} \frac{\hbar^2}{m\pi^2} \left(3\pi^2 \frac{N}{V} \right)^{5/3}.$$

4.10. Consider a system with the band dispersion

$$\varepsilon(k) = -[E_1 \cos(k_x a) + E_2 \cos(k_y a) + E_3 \cos(k_z a)],$$

as calculated within tight-binding approximation. The constants E_1 , E_2 , and E_3 are positive. Show that the specific heat of almost empty or almost filled bands is the one of a gas of free particles.

Solution. If the band is almost empty, $\varepsilon(k)$ can be expanded in the vicinity of $k = 0$, obtaining the principal effective masses

$$m_i = \frac{\hbar^2}{E_i a^2}, \quad i = 1, 2, 3. \quad (4.236)$$

Notice that $m_1 m_2 m_3 = \det(M)$, with M the mass tensor. The density of states for such a system has been calculated in the main text. There, it has been found that the functional dependence on energy is the same in both cases of isotropic and anisotropic masses, except that $m^{\frac{3}{2}}$ in the former case is to be substituted with $(m_1 m_2 m_3)^{1/3}$ in the latter. Therefore, one expects that the behavior of the specific heat with respect to temperature remains unchanged, except for a numerical factor.

Since the band maximum is at $(\pi/a, \pi/a, \pi/a)$, following the same procedure as in the isotropic case one finds the same effective masses except for a change in sign. On the other hand, the density of states enters $(|m_1 m_2 m_3|)^{1/3}$, and therefore

the results are the same. On physical grounds, an almost filled band is expected to have a vanishing specific heat because thermal excitations of electrons are highly hampered by the Pauli exclusion principle.

4.11. Show that the resistivity tensor has a positive determinant.

Solution. The resistivity tensor $\hat{\rho}$ is defined by

$$\mathbf{E} = \hat{\rho} \cdot \mathbf{j}.$$

The principal components of the tensor are ρ_{xx} , ρ_{yy} , ρ_{zz} , so that the dissipated power is $P = \mathbf{E} \cdot \mathbf{j} = \rho_{xx} j_x^2 + \rho_{yy} j_y^2 + \rho_{zz} j_z^2$. The latter is to be positive, where $\rho_{ii} > 0$ for each i results from. On the other hand, the determinant is $\det(\hat{\rho}) = \rho_{xx}\rho_{yy}\rho_{zz} > 0$. Since the determinant does not depend on the reference system with respect to which it is calculated, the problem statement follows.

4.12. The magneto-conductivity tensor in a material with electrons and holes as charge carriers has been calculated in the main text, along with the Hall coefficient in various physical situations. Calculate the resistivity difference $\Delta\rho = \rho(B) - \rho(0)$ and the Hall coefficient.

Solution. In the main text it has been shown that the total current can be cast in the form

$$\mathbf{j} = (\hat{\sigma}_1 + \hat{\sigma}_2) \mathbf{E}.$$

Thus, the conductivity tensor is given by (4.203) with components:

$$\begin{aligned} \sigma_{xx} = \sigma_{yy} &= \frac{\sigma_1}{P_1} + \frac{\sigma_2}{P_2} \\ \sigma_{xy} = -\sigma_{yx} &= -\frac{\sigma_1 \omega_{1c} \tau_1}{P_1} + \frac{\sigma_2 \omega_{2c} \tau_2}{P_2}. \end{aligned} \quad (4.237)$$

One then finds

$$\begin{aligned} \sigma_{xx} &= e^2 \left(\frac{n_1 \mu_1}{P_1} + \frac{n_2 \mu_2}{P_2} \right) \\ \sigma_{xy} &= e^2 \left(-\frac{n_1 \mu_1 \omega_{1c} \tau_1}{P_1} + \frac{n_2 \mu_2 \omega_{2c} \tau_2}{P_2} \right), \end{aligned} \quad (4.238)$$

with $\mu_1 = \tau_1/m_1$ and $\mu_2 = \tau_2/m_2$. In addition, one has

$$\begin{aligned} \sigma_{xx}^2 + \sigma_{xy}^2 &= e^4 \left[\frac{(n_1 \mu_1)^2}{P_1} + \frac{(n_2 \mu_2)^2}{P_2} \right. \\ &\quad \left. + \frac{2n_1 \mu_1 n_2 \mu_2}{P_1 P_2} (1 - \omega_{1c} \tau_1 \omega_{2c} \tau_2) \right]. \end{aligned} \quad (4.239)$$

At zero current $j_y = 0$, it results that $E_y = \sigma_{xy} E_x / \sigma_{xx}$ and $j_x = (\sigma_{xx} + \sigma_{xy}^2 / \sigma_{xx}) E_x$. Under such conditions, the magneto-resistivity along the x-axis is $\Delta\rho = \rho(B) - \rho(0)$ and the ratio $\Delta\rho/\rho(0)$ is

$$\begin{aligned}\frac{\Delta\rho}{\rho(0)} &= \left[\frac{1}{\sigma_{xx} + \frac{\sigma_{xy}^2}{\sigma_{xx}}} - \frac{1}{\sigma_{xx}} \right] \sigma_{xx} \\ &= -\frac{\sigma_{xy}^2}{\sigma_{xx}^2 + \sigma_{xy}^2}.\end{aligned}\quad (4.240)$$

Two limiting cases can be studied:

(a) Case with $\omega_{1c}\tau_1, \omega_{2c}\tau_2 \ll 1$. Since $P_1 = 1$ e $P_2 = 1$, one easily finds that

$$\begin{aligned}\sigma_{xy} &= e^2 (-n_1\mu_1\omega_{1c}\tau_1 + n_2\mu_2\omega_{2c}\tau_2) \\ &= \frac{e^2}{c^2} (n_2\mu_2 - n_1\mu_1), \quad \sigma_{xx}^2 + \sigma_{xy}^2 = (n_1\mu_1 + n_2\mu_2)^2,\end{aligned}\quad (4.241)$$

and

$$\frac{\Delta\rho}{\rho(0)} = -\frac{e^2}{c^2} \frac{(n_2\mu_2^2 - n_1\mu_1^2)^2}{(n_1\mu_1 + n_2\mu_2)^2} B^2.$$

In particular, if in addition $n_1 = n_2 = n$, one obtains

$$\frac{\Delta\rho}{\rho(0)} = -\frac{e^2}{c^2} (\mu_2 - \mu_1)^2 B^2.$$

If instead $\mu_1 = \mu_2 = \mu$, one has

$$\Delta\rho = -\frac{e^2}{c^2} \mu^2 B^2 \frac{(n_2 - n_1)^2}{(n_2 + n_1)^2}.$$

(b) Case with $\omega_{1c}\tau_1, \omega_{2c}\tau_2 \gg 1$. Since

$$\begin{aligned}\sigma_{xy} &= e^2 \left(-\frac{n_1\mu_1}{\omega_{1c}\tau_1} + \frac{n_2\mu_2}{\omega_{2c}\tau_2} \right) \\ , \sigma_{xx}^2 + \sigma_{xy}^2 &= e^4 \left[\left(\frac{n_1\mu_1}{\omega_{1c}\tau_1} \right)^2 + \left(\frac{n_2\mu_2}{\omega_{2c}\tau_2} \right)^2 - \frac{2n_1\mu_1 n_2\mu_2}{\omega_{1c}\tau_1 \omega_{2c}\tau_2} \right],\end{aligned}\quad (4.242)$$

one finds that

$$\frac{\Delta\rho}{\rho(0)} = 1.$$

Problems without solutions

4.13. Consider the one-dimensional band $E(k) = E_0 + E_1 \cos(ka)$. Calculate the group velocity of the electron as a function of time, after the application of a constant electric field.

4.14. Consider the two-dimensional electronic band $E(k_x, k_y) = E_0 + E_1(\cos(k_x a) + \cos(k_y a))$, corresponding to a square lattice with side a . The components of \mathbf{E} are $\mathbf{E} \equiv (E_0, E_0, 0)$. Calculate the components of the group velocity and the frequency of the Bloch oscillations, if the electron is in the initial state with $\mathbf{k}_0 = 0$. Repeat the calculation, after changing the electric field components into $\mathbf{E} \equiv (E_0, E_0/2, 0)$.

4.15. Consider the electronic band given in Problem 4.14. Find first the critical points. Then, place a magnetic field perpendicular to the square lattice, and determine the conditions for which the trajectories with energy close to these points are closed or open.

4.16. Consider the electronic band given in Problem 4.14. Determine the hole effective masses close to the critical points.

4.17. Consider an electron gas with density n , electron-electron scattering cross section $\sigma_u = 10^{-16} \text{ cm}^2$, and temperature T . Calculate the mean free path with and without accounting for the Pauli exclusion principle.

4.18. Consider the Landau states of a two-dimensional electron gas. Calculate the dipole matrix elements in the plane perpendicular to the magnetic field.

References

1. L. Bloomfield: *How Everything Works. Making physics out of ordinary*. John Wiley and Sons, USA (2008)
2. N.W. Ashcroft and N. D. Mermin: *Solid-State Physics*. 2nd ed. Holt-Saunders, Tokyo (1976)
3. F. Bassani and U.M. Grassano: *Fisica dello Stato Solido*. 1st ed. Bollati Boringhieri, Torino (2000)
4. S. Datta: *Quantum transport*. 1st ed. Cambridge Univ. Press (2005)
5. J. M. Ziman: *Principles of the theory of solids*. 2nd ed. Cambridge University Press, Cambridge (1964)
6. M. Dahan, E. Peik, J. Reichel, Y. Castin, and C. Salomon, *Phys. Rev. Lett.* **76**, 4508 (1996)
7. For a review, see e.g. E. E. Mendez and G. Bastard, *Physics Today* **46**, 34 (1993)
8. For a review on more recent methods, see e.g. A. Alberti, G. Ferrari, V. V. Ivanov, M. L. Chiofalo, and G. Tino, *New J. of Phys.* **12**, 65037 (2010)
9. N. Poli, F.-Y. Wang, M. G. Tarallo, A. Alberti, M. Prevedelli, and G. M. Tino, *Phys. Rev. Lett.* **106**, 038501 (2011)
10. Data and figure arrangement are courtesy of Marco Tarallo [11]
11. M. G. Tarallo, A. Alberti, N. Poli, M. L. Chiofalo, F.-Y. Wang, and G. M. Tino, *Phys. Rev. A* **86**, 033615 (2012)
12. C. Kittel: *Introduction to Solid State Physics*. 6th ed. Wiley, New York (1986)
13. M. Alvisi and A. Licciulli: *Le proprietà elettriche dei materiali*, Corso di Scienza ed Ingegneria dei materiali, <http://www.antonio.licciulli.unisalento.it/didattica2006/proprietielettriche.pdf> (2006)
14. R. Berman and D. K. C. MacDonald, *Proc. Roy. Soc. A* **209**, 368 (1952)
15. W.H. Lien and N.E. Phillips, *Phys. Rev.* **133**, A1370 (1964)
16. W.S. Corax, M.P. Garfunkel, C.B. Satterhwaite, and A. Wexler, *Phys. Rev.* **98**, 1699 (1955)
17. A. Alessandrello, D.V. Camin, E. Fiorini, A. Giuliani, M. Buraschi, and G. Pignatelli, *Europhys. Lett.* **7**, 69 (1988)

18. Hai Yang Bai, Jim Lin Luo, Duo Jin, and Ji Rong Sun, *J. Appl. Phys.* **79**, 361 (1996)
19. R. Luck, *Phys. Stat. Sol.* **18**, 49 (1966)
20. H.J. Fritzsche, *J. Phys. Chem. Solids* **6**, 69 (1958)
21. G. Burns: *Solid State Physics*. 1st ed. Academic Press Inc., Orlando (1985)
22. R. Berman, F.S. Simons, and J.M. Ziman, *Proc. Royal Soc. London* **A220**, 171 (1953)
23. K. von Klitzing, *Europhysics News* **13**, 2 (1982). See also the original papers K. von Klitzing, G. Dorda, and M. Pepper, *Phys. Rev. Lett.* **45**, 494 (1980); R. B. Laughlin. *Phys. Rev. B* **23**, 5632 (1981)

Chapter 5

Optical Properties

Abstract Elementary optical properties of crystals and nanostructures are investigated, like absorption, emission and reflectivity, as the system response to time- and space-dependent electric fields. The corresponding response and dielectric functions are introduced, that are dispersive in frequency and wavenumber. Simple models of dielectric functions for metals and insulators are introduced and studied. Electronic transitions induced by electromagnetic fields assisted by phonon excitations are investigated as well, and their occurrence determined to first and second perturbative order. Finally the optical properties of confined nanostructures are discussed. The issue of electronic and ionic screening of interactions is considered for all the cases under discussion.

5.1 Introduction

Did you ever ask how microwave oven work to heat up and cook food, it how happens that lasers and LEDs produce such bright lights, how it occurs that some materials used everyday are reflectant while others are opaque to light, as it occurs for example with paints and makeups [1]. The content of this chapter may open up useful answers [1].

After investigating the charge and energy transport behavior in response to external static electric, magnetic, and temperature fields, the response of infinite crystals to external electromagnetic fields is indeed addressed in the present Chapter. In particular, the optical properties of metals, semiconductors and insulators are investigated in regard to two characterizing phenomena: propagation of electromagnetic radiation in materials and coefficients of optical emission and absorption. The observed quantities are macroscopic in character, so that the theoretical framework might to first approximation effectively remain either classical or quantum. However, selected phenomena are originated by purely quantum effects, and are consequently treated in a quantum manner [2, 3, 4, 5].

Propagation and absorption or emission of radiation are related phenomena within the linear response approach here outlined: they are in fact represented by the real and imaginary parts of one and the same function, that is the dielectric function. The latter depends on both frequency and wavevector. The wavevector dependence sets the existence of regions where relevant screening effects of Coulomb-like interactions occur, while they are entangled with propagation and absorption or emission phenomena. Correlations and exchange processes change the charge density distribution around a given particle: in fact, the system behaves as if the real particles were substituted by larger particles with size on the spatial scale of the density change, interacting through short- range interactions. At vanishing frequency, propagation and screening decouple and can be treated independently of each other.

A second connection level can be actually set. Transport behavior and optical properties of materials are both dictated by the way a material responds to suited external disturbances. This response tells the internal structure of the material and its characterizing interactions, since the power dissipated by the external disturbance is related to the way the induced fluctuations decay once they are created. For small perturbation amplitudes, when a theory of the linear response is suited, a tight connection exists between the measurement of the low-frequency, down to static, transport coefficients investigated in Chap. 4 and that of the dielectric and magnetic susceptibilities investigated in this Chapter along with the optical conductivity in the same low-frequency regime. In fact, the two are sides of the same coin: transport coefficients are related to long-wavelength first and then low-frequency limiting behaviors of current response functions, whereas static susceptibilities to low-frequency first and then long-wavelength limits of density response functions. In the low-frequency regime, linear i.e. small-amplitude response theory reduces to linearized hydrodynamic theory and viceversa. The linear response-function theory approach is most useful to treat slightly perturbed frequency-dependent phenomena, while hydrodynamics is better suited to treat even wide amplitude but low-frequency perturbations.

On the other hand, while reflecting the system inhomogeneity and embodying the correlations between the system particles, current and density responses are related to each other via general conservation laws for density, momentum, and energy, as well as to thermodynamic relations [6]. All these pieces of knowledge that are built up in Chap. 4 and in this Chapter, are generalized and assembled together in Chap. 6 under general conditions within the framework of Density and Current-Density Functional Theory. This provides an elegant approach where all these concepts are consistently presented, along with a conceptual and factual route to determine the properties of a strongly correlated and slightly inhomogeneous system in terms of a one-to-one mapping to a fictitious system that is non-interacting and homogeneous.

The special confinement-related properties of nanostructures have striking effects on their optical properties: the case of quantum wells and quantum wires is considered in a final part of the Chapter. Further discussions on this topic with a specific example on quantum dots, is postponed to the end of Chap. 6 as an application of the

Density Functional Theory. The Chapter ends with a discussion on the experimental methods suited to determine the optical properties of crystals and nanostructures.

5.2 Response to electromagnetic fields

The basic tools are here introduced, to investigate the properties of electrons and ions as they manifest in response to external electromagnetic disturbances. At variance with the static probes considered in Chap. 4, the electric and magnetic fields considered here originate disturbances that are rapidly varying in both space and time. The focus of this section is on the system behavior in regard to radiation propagation, absorption and emission, and their interplay.

Most of the properties can be described by macroscopic quantities and the classical physics framework is enough. Thus, a simple phenomenological classical model is first introduced, to effectively catch the relevant conceptual characteristics of the response of charged particles to internal forces and external electromagnetic fields [6]. The solution to this model is then analyzed in regard to its general properties. The model is eventually applied to different physical situations in comparison with experiments. The introduction of a quantum physics framework is the final step.

5.2.1 A simple useful model

The simple example model is that of a damped harmonic oscillator subjected to an external force: the choice is grounded on the idea that excitations in matter often behave as if they were sets of damped harmonic oscillators. Real manifestations of such an idea have been continuously at hand in the previous Chapters.

Definition of the model. Consider thus the following equation of motion for a single particle with mass m moving in one dimension [6]:

$$\ddot{y}(x,t) + \frac{\dot{y}(x,t)}{\tau} + \omega_0^2 y(x,t) = \frac{F(x,t)}{m}, \quad (5.1)$$

where $y(x,t)$ is the particle displacement at position x and time t out of its equilibrium position, induced by the action of the space and time dependent external force $F(x,t)$. The force may induce an oscillatory motion, that is here modeled by means of an elastic force characterized by a natural frequency ω_0 . The force may also induce a current, that is the particle may also be drifted while suffering dissipative internal interactions, due e.g. to various friction and scattering mechanisms as those analyzed in Chap. 4. The latter are modeled by a viscous force with a characteristic decay time τ or damping coefficient $\gamma = (2\tau)^{-1}$.

In the absence of the external force, it is well known from classical mechanics that the solution of (5.1) is a damped harmonic oscillator: the particle motion as represented by the coordinate $y(x, t)$ would execute oscillations in time at the natural frequency ω_0 , the initial oscillation amplitude would be exponentially damped in time. In the absence of the external force, the average value at equilibrium of each of the terms on the left-hand-side of (5.1) would indeed vanish. This is not the case when the external force drives the system out of equilibrium: on average the non-equilibrium displacement $w_{ne}(x, t) \equiv \langle y(x, t) \rangle_{ne}$, as calculated with respect to the equilibrium ensemble. If $F(x, t) \propto \delta(t - t')$ were a sharp δ -like impulse impressed at a given time t' , the solution of (5.1) can be found after Fourier transforming it side-by-side:

$$w_{ne}(x, \omega) \propto \frac{1}{m} \frac{1}{(\omega_0^2 - \omega^2) - 2i\gamma\omega}. \quad (5.2)$$

In time Fourier space, this corresponds to damped harmonic-oscillator solution $w_{ne}(x, t) \propto \omega_\gamma^{-1} e^{-\gamma t} \sin(\omega_\gamma t) \theta(t - t')$ with $\omega_\gamma \equiv \sqrt{\omega_0^2 - \gamma^2}$ and $\theta(t - t')$ the Heaviside step function.

Definition of the response function. The action of a generic time- and space-dependent force can be viewed as a weighted sum of an infinite number of such impulses, so that the average non-equilibrium displacement can be formally connected to the external force by means of the relation

$$w_{ne}(x, t) = \int_{-\infty}^{+\infty} dx' \int_{-\infty}^{+\infty} dt' \chi(x, x'; t, t') F(x', t') \theta(t - t'). \quad (5.3)$$

Definition

Expression (5.3) defines the response function $\chi(x, x'; t, t') \theta(t - t')$.

In fact, (5.3) expresses the idea that the displacement at position (x, t) is determined by the response behavior of the system in the whole remaining space x' to a unit impulsive external force applied at an earlier time t' . Indeed, the response cannot change after a temporal translation of both t and t' by the same Δt . Therefore, the response χ is to be a function of $t - t'$, that is $\chi(x, x', t - t')$. Because of the causality principle built in, $\chi(x, x'; t - t') \theta(t - t')$ is called retarded response function, and it actually works as a Green's function for the system in the sense discussed above in the simple example. As to the space dependence of χ instead, some more elaboration is needed. Two different length scales characterize (5.1): the average particle displacement $\text{Re}(w_{ne}(x, t))$ and the wavelength Λ of the external field. The external position-dependent force determines a spatial inhomogeneity. However, in the limiting regime when $|w_{ne}(x, t)| \ll \Lambda$ and if the system is homogeneous over the length $\lambda \ll \Lambda$, the field can be considered to be constant along the whole displacement: as a result, $\chi(x, x'; t, t') = \chi(x - x'; t - t')$ can be viewed as a function of $x - x'$. These are

common enough conditions. If not, all the reasoning below can be easily extended at the price of complicating the equations.

In the following, the seminal properties of the response function are derived for the simple model of the damped harmonic oscillator and under the above conditions of space and time translational invariance. It is demonstrated later on in Chapter that these properties apply to the general case. Going back to (5.1), the solution is easily found after operating a Fourier transform in both space and time variables. One has indeed

$$w_{ne}(q, \omega) = \frac{F(q, \omega)}{m} \frac{1}{(\omega_0^2 - \omega^2) - i\gamma\omega}. \quad (5.4)$$

In order to keep the notation simple, the q -dependence can be omitted and the same symbols w_{ne} and F have been used to represent the displacement and external force in both time and frequency domain. In particular when $|w_{ne}(x, t)| \ll \Lambda$ it can be shown that $q \rightarrow 0$ in (5.4), as it is shown in Appendix 5.7. Fourier transform of (5.3) and comparison with (5.4) lead to the expression for the response function of the damped harmonic oscillator:

$$\chi(\omega) = \frac{1}{m} \frac{1}{-\omega^2 + \omega_0^2 - i\gamma\omega}. \quad (5.5)$$

Real and imaginary part of the response function. The response function is usually separated into its real $\chi'(\omega)$ and imaginary $\chi''(\omega)$ parts, that are

$$\begin{aligned} \chi'(\omega) &= \frac{1}{m} \frac{\omega_0^2 - \omega^2}{(\omega^2 - \omega_0^2)^2 + (\gamma\omega)^2} \\ \chi''(\omega) &= \frac{1}{m} \frac{\gamma\omega}{(\omega^2 - \omega_0^2)^2 + (\gamma\omega)^2}. \end{aligned} \quad (5.6)$$

Notice that χ' is an even and χ'' an odd function of ω . In particular, $\chi'(\omega)$ has the typical shape of a dispersive spectrum as in Fig. 5.1 (a), while $\chi''(\omega)$ the typical shape of an absorption spectrum as in Fig. 5.1 (b). Through the parameters ω_0 and γ or τ , the response function contains information on the internal modes and their damping processes due to interactions within the system. This has crucial consequences on the physical meaning of the real and imaginary parts of $\chi(\omega)$ as the dispersive and absorptive response functions. In the following this concept is explored.

The fluctuation-dissipation theorem. The absorptive and dissipative role of $\chi''(\omega)$ neatly emerges when one calculates the power which is dissipated over one period of a single-frequency external force $F(t) = \text{Re}[F_0 e^{i\omega t}] \equiv \text{Re}[F_\omega]$. This is

$$W(\omega) = (2\pi/\omega)^{-1} \int_0^{2\pi/\omega} dt \mathbf{F}(t) \cdot \mathbf{v}(t),$$

in terms of the average particle velocity $\mathbf{v}(t)$. In the present simple 1D model, $\mathbf{F}(t) \cdot \mathbf{v}(t) = F(t)v(t)$. Using the properties of the derivatives of Fourier transforms, one

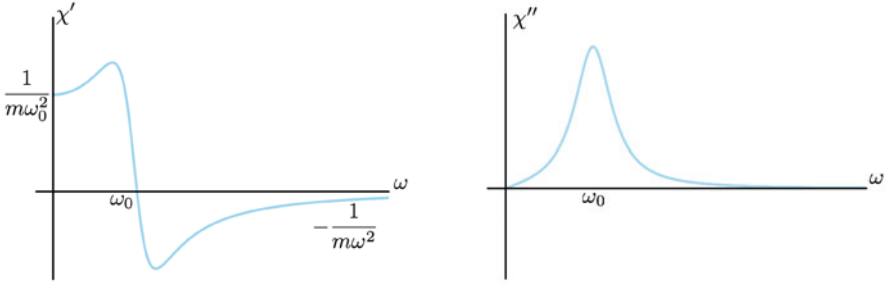


Fig. 5.1 Simple model for the response function. The typical real and imaginary parts of $\chi(\omega)$ are sketched. Left panel: the real part, evidencing the typical dispersive shape and the $\omega \rightarrow 0$ and $\omega \rightarrow \infty$ behaviors as described in main text. Right panel: the imaginary part, evidencing the typical absorptive lorentzian shape

has $\text{Re}[\dot{w}_{ne}(t)] = \text{Re}[i\omega w_{ne}(\omega)] = \text{Re}[i\omega\chi(\omega)F_\omega]$. Therefore, one obtains

$$W(\omega) = \frac{1}{2\pi/\omega} \int_0^{2\pi/\omega} dt \frac{F_0 e^{i\omega t} + F_0^* e^{-i\omega t}}{2} \frac{i\omega\chi(\omega)F_0 e^{i\omega t} - i\omega\chi^*(\omega)F_0^* e^{-i\omega t}}{2}.$$

The terms in $e^{\pm 2i\omega t}$ average to zero over one period. Only the time-independent terms, where $e^{i\omega t}$ and $e^{-i\omega t}$ cancel out, are eventually left, so that: $W(\omega) = i\omega|F_0|^2(\chi(\omega) - \chi^*(\omega))/4 = -\omega\chi''(\omega)|F_0|^2/2$. This relation can be cast in the more expressive form:

$$W(\omega) = -\omega\chi''(\omega)|F(t)|_{\text{ave}}^2, \quad (5.7)$$

that is in terms of the average quadratic force $|F(t)|_{\text{ave}}^2$ over one period. In essence, (5.7) states that

Concept

Over one period of the external disturbance, the average power dissipated is connected to the average squared external force acting on the medium: the link is provided by the imaginary part of the response function. Therefore, the imaginay part of the response function provides information on frequencies and damping of the system modes, or else of its excitations.

In practice, the external field acts to resonantly drive natural excitation modes of the system: once the external field is removed, a look at the way in which these excitations are damped out provides information on ω_0 and τ and thus on the excitation frequency and on the interaction processes. Though derived within a simplest classical model for a damped and forced harmonic oscillator, expression (5.7) is one form of the very general so-called fluctuation-dissipation theorem. The statement can indeed be generalized to a set of N independent harmonic oscillators and it can also be extended to quantum systems of many interacting particles, as enlight-

ened in (5.67) later on within a more refined quantum treatment. The fluctuation-dissipation theorem simplified into (5.7) is therefore of crucial importance, since it sets the conceptual and practical relationship between system fluctuations and dissipation. Measurements are based on this concept, such as those described at the end of Chap. 1-4.

Limiting behavior of the response function and sum rules. As to the real part of the frequency-dependent response function, notice that in the low frequency limit $\chi'(\omega \rightarrow 0) = (m\omega_0^2)^{-1}$: the static response implies no dissipation, since a time-independent force determines a time-independent displacement and produces no dissipation. On the other hand, in the high-frequency limit $\chi'(\omega \rightarrow +\infty) = -(m\omega^2)^{-1}$, not depending on γ or ω_0 : the oscillator responds out-of-phase with respect to the external force and independently of the environment. Thus, dissipation washes away at low and high frequencies. Indeed, by inspection the imaginary response functions in the same limits vanish: $\chi''(\omega) \rightarrow 0$ when $\omega \rightarrow 0$ and $\chi''(\omega) \rightarrow 0$ when $\omega \rightarrow \infty$.

Finally, consider the absorptive function $\chi''(\omega)$ in (5.6): its frequency average $\int_{-\infty}^{+\infty} \pi^{-1} d\omega \chi''(\omega) \omega = 1/m$ is independent of the details of the internal interaction among the particles. Therefore, if N independent oscillators are considered, one may state that such quantity counts the number of particles in the system. Irrespectively of its simplicity, this is a general result demonstrated in Appendix 5.10, expressing an integral condition, or so-called sum rule, that the exact imaginary part of the response function has to fulfill. These sum rules are quite useful tools to validate approximated models for the response function.

The simple and meaningful model illustrated here [6] is going to be used throughout the present and the next Chapters.

Case of an external electric field . The above findings can be easily specialized to the purposes of the present Chapter. In this case, the external force is originated by an electric field through $F(x, t) = eE_{\text{ext}}(x, t)$ acting on the charge e of the particle. A few specific notes are in order. First, the electric force can also induce a polarization of the medium: in the model, this is accounted for by intending that the total electric force is originated by the external electric field and by the induced field inside the material. Second, the magnetic part of the electromagnetic field is conveniently neglected, as the charged particle is assumed to acquire much smaller velocities than the speed of light. Finally, the nature of the charged particles is not specified yet, they being either electrons or ions, or else. In any event, the assumed quasi-homogeneity condition of the system is guaranteed by asking that the mean-free path λ of the electrons be much smaller than Λ , that is $\lambda \ll \Lambda$.

Frequency-dependent polarization. An observable is now being introduced, that is relevant in the context of the present Chapter. In fact, in response to the electric field the system polarizes. Therefore, the suited observable to look at is represented by the polarization $P(\omega) = Nw_{ne}(\omega)$, that is proportional to the displacement $w_{ne}(\omega)$ through the number N of particles, each with charge e . The corresponding response can be represented by the dielectric susceptibility $\chi(\omega) \equiv P(\omega)/E_{\text{ext}}(\omega)$:

$$\chi(\omega) = \frac{Ne^2}{m} \frac{1}{\omega_0^2 - \omega^2 - i\gamma\omega}. \quad (5.8)$$

Once again, the structure of both the real and imaginary parts of (5.8) reveals that the response of the system is significant in correspondence of its natural oscillation frequency ω_0 . A more precise and quantitative account of this statement is provided later on in the text.

Frequency-dependent dielectric function and relationship with the response function. Since the total macroscopic electric field $\mathbf{E}(\omega) = \mathbf{E}_{\text{ext}}(\omega) + \mathbf{E}_{\text{ind}}(\omega)$ is composed by the field $\mathbf{E}_{\text{ext}}(\omega)$ in vacuum and the induced part $\mathbf{E}_{\text{ind}}(\omega) = -4\pi\mathbf{P}(\omega)$, a new quantity is conveniently introduced as detailed in Appendix 5.7, that is the dielectric function.

Concept

The dielectric function $\varepsilon(\omega) = E_{\text{ext}}(\omega)/E(\omega)$, essentially measures the capacity of the medium of screening the acting external field.

The ratio $E_{\text{ext}}(\omega)/E(\omega)$ above is intended to be between two same components $E_{\text{ext},i}(\omega)$ of the vectors \mathbf{E}_{ext} and \mathbf{E} , that is $\varepsilon(\omega)E_i(\omega) = E_{\text{ext},i}(\omega)$. The definition of response function $P(\omega) = \chi(\omega)E_{\text{ext}}(\omega)$ provides the link between dielectric and response functions, that is

$$\frac{1}{\varepsilon(\omega)} = \frac{E(\omega)}{E_{\text{ext}}(\omega)} = 1 - 4\pi\chi(\omega). \quad (5.9)$$

To the first order it is found

$$\varepsilon(\omega) = 1 + 4\pi\chi(\omega). \quad (5.10)$$

Notice that the starting expression (5.3) connects the displacement to the external force $F(x, t)$. An alternative definition is the same relation (5.3), but with $F(x, t)$ indicating the total force acting on the particle, i.e. $F(x, t)$ is the force due to the external one plus that due to the induced effects on the material. The analogous definition for the response function can be introduced, and the connection between the polarization and the total electric field becomes $\chi_0(\omega) \equiv P(\omega)/E(\omega)$. Finally the definition of the dielectric function becomes

$$\varepsilon(\omega) = \frac{E_{\text{ext}}(\omega)}{E(\omega)} = \frac{E(\omega)(1 + 4\pi\chi_0(\omega))}{E(\omega)} = 1 + 4\pi\chi_0(\omega). \quad (5.11)$$

The quantities $\chi_0(\omega)$ and $\chi(\omega)$ differ for next-to-linear order terms. Thus, to linear order the response of the system to the external and to the total field do coincide.

Reverting back to (5.8), one then finds that

$$\varepsilon(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\omega/\tau}. \quad (5.12)$$

Here, ω_p is a characteristic frequency given by $\omega_p^2 = 4\pi Ne^2/m$: it is named plasma frequency and represents the frequency of the long-wavelength collective oscillation mode of the charged particles in the system, as e.g. are ions, electrons, holes. The dielectric function (5.12) is thus a complex function of the real variable ω , whose real and imaginary parts are

$$\begin{aligned}\epsilon_1(\omega) &= 1 + \omega_p^2 \frac{(\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2 + \omega^2/\tau^2} \\ \epsilon_2(\omega) &= \omega_p^2 \frac{\omega/\tau}{(\omega_0^2 - \omega^2)^2 + \omega^2/\tau^2}.\end{aligned}\quad (5.13)$$

Frequency-dependent conductivity. Up to this point, the polarization response of the system has been considered. It can be anticipated that actually the complex frequency-dependent dielectric function (5.12) accounts for the system capacity to both polarize and conduct, that is to respond to the external field by inducing polarization phenomena and currents. The real and imaginary parts (5.13) of the dielectric function respectively drive the phenomena of radiation propagation and of absorption. Appendix 5.7 details the general definitions and properties of \mathbf{q} and ω dependent dielectric functions, and details the longitudinal and transverse solutions of the Maxwell equations.

Quick Questions

Q1. In the simple model used so far, $\chi''(\omega)$ has been shown to be an odd function of ω . This is in fact a more general statement, independent of the specific model. Provide a simple convincing argument [Hint: use the concept of power dissipation].

Answer - $W(\omega)$ is to be a positive quantity. Thus $\omega\chi''(\omega) \leq 0$ at all frequencies. Therefore, it must be $\chi''(-\omega) = -\chi''(\omega)$.

Q2. Consider a system that is first driven by an external force at resonance frequency, say $\omega = \omega_0$ and then at very low $\omega \rightarrow 0$ frequency. Which one between the two driving has deeper effect on the propagation properties? Which one is more efficient in reducing absorption? .

Answer - Driving at resonance makes $\epsilon_1(\omega_0) \rightarrow 1$ and $\epsilon_2(\omega_0) \rightarrow \omega_p^2\tau/\omega_0$; thus, it favors absorption. Driving at $\omega \rightarrow 0$ makes $\epsilon_1(\omega \rightarrow 0) \rightarrow 1 + \omega_p^2/\omega_0^2$ and $\epsilon_2(\omega \rightarrow 0) \rightarrow 0$; thus, it affects the propagation and depresses the absorption.

5.2.2 Dielectric function, polarizability and conductivity

5.2.2.1 Dielectric function and polarizability

It is known that electromagnetic radiation traveling between different media may determine the occurrence of polarization and currents. Radiation impinging on a medium may propagate or be absorbed, and again undergo phenomena of refraction, reflection and transmission. In particular, refraction is a consequence of the different speed which the light travels with in different materials. The question arises on how the dielectric function formulation is connected to other physical optics quantities that are met in everyday life. This is the aim of this section.

In regard to the occurrence of a system polarization, from elementary physics it is known that propagation of electromagnetic radiation in a medium with refraction index n is characterized by the dispersion law $k = \omega n/c$. It is also known that the refraction index may depend on the radiation frequency. In this case, a driving plane-wave electric field $\mathbf{E}(x, t)$ that propagates along the x axis can be cast in either one of the equivalent Fourier transform forms

$$\begin{aligned}\mathbf{E}(x, t) &= \int \mathbf{E}(\omega) e^{i(k(\omega)x - \omega t)} d\omega \\ \mathbf{E}(x, t) &= \int \mathbf{E}(k) e^{i(kx - \omega(k)t)} dk.\end{aligned}\tag{5.14}$$

The functions $k(\omega)$ or $\omega(k)$ determine the following properties.

Properties

P1. Case with $k(\omega)$ and $\omega(k)$ real functions. The electric field propagates into the material in a dispersive manner: each frequency component has a different phase velocity and the group velocity is given by $d\omega(k)/dk$.

P2. Case with $k(\omega) = k_1(\omega) + ik_2(\omega)$ a complex function of the real variable ω . The imaginary part of $k(\omega)$ in the first line of (5.14) drives the amplitude of the electric field to be exponentially depressed or enhanced with distance, whereas the real part drives the propagation behavior.

P3. Case with $\omega(k) = \omega_1(k) + i\omega_2(k)$ a complex function of the real variable k . The imaginary part of $\omega(k)$ in the second line of (5.14) drives the amplitude of the electric field to be exponentially depressed or enhanced with time, whereas the real part drives the propagation behavior.

The complex frequency-dependent refraction index $\tilde{n} \equiv n + i\kappa$ is conveniently introduced at this point: \tilde{n} is connected to the polarization part of the dielectric

function $\varepsilon(\omega)$ by the relation

$$\tilde{n}(\omega) = n(\omega) + i\kappa(\omega) = \sqrt{\varepsilon(\omega)}. \quad (5.15)$$

One then has

$$(n + i\kappa)^2 = \varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$$

and therefore $n^2 - \kappa^2 = \varepsilon_1(\omega)$, $2n\kappa = \varepsilon_2(\omega)$. Neglecting terms of order $(\kappa/n)^2$, one obtains $k_1 = \omega n/c$ and $k_2 = \omega \kappa/c$. Since the radiation intensity I is $\propto |E|^2$, one has $I = I_0 e^{-\eta x}$.

Definition

The relation $I = I_0 e^{-\eta x}$ defines the absorption coefficient $\eta = 2\omega\kappa/c = \omega\varepsilon_2/(nc)$

Concept

In essence, the real part of the dielectric function drives the propagation of radiation and the imaginary part determines the absorption phenomena.

In the simple model (5.4)-(5.8), the polarization is the manifestation of the system linear response to the application of an external electric field. Let $\chi_p(\mathbf{r}, \mathbf{r}', t, t')$ be the response function to the total electric field $\mathbf{E}(\omega)$. The general definition of polarization $\mathbf{P}(\mathbf{r}, t)$ after the application of a space- and time-dependent electric field is:

$$\mathbf{P}(\mathbf{r}, t) = \int d\mathbf{r}' \int_{-\infty}^t dt \chi_p(\mathbf{r}, \mathbf{r}', t - t') \mathbf{E}(\mathbf{r}', t'). \quad (5.16)$$

As already observed in the simple model, the expression above reveals that the polarization at position and time (\mathbf{r}, t) is determined by the response behavior of the system in the whole remaining space \mathbf{r}' and at all earlier times t' . This observation has crucial consequences on the frequency behavior of the system response, that are outlined below.

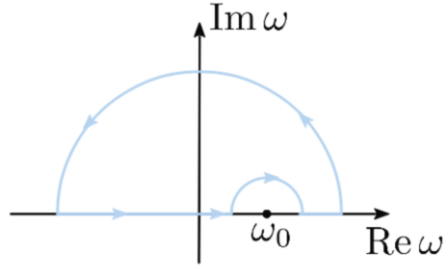
In order to ease the notation, the space dependence of χ_p is in the following omitted. In response to the component $\mathbf{E}_0 e^{-i\omega t'}$ of $\mathbf{E}(t')$, after a change of variables the polarization can be cast into the form

$$\mathbf{P}(t) = \int_{-\infty}^t dt \chi_p(t - t') \mathbf{E}_0 e^{-i\omega t'} = \mathbf{E}_0 e^{-i\omega t} \int_0^{\infty} d\tau \chi_p(\tau) e^{i\omega \tau}. \quad (5.17)$$

This form remarks that the time dependence of $\mathbf{P}(t)$ is dictated by the total electric field. The properties of the frequency component of \mathbf{P} are instead embodied in the time integral:

$$\alpha(\omega) = \int_0^{\infty} d\tau \chi_p(\tau) e^{i\omega \tau}. \quad (5.18)$$

Fig. 5.2 Polarization and dielectric function. Sketch of the contour used in the complex plane to integrate the function $\alpha(\omega)/(\omega - \omega_0)$



The function $\alpha(\omega)$ represents the dielectric susceptibility $\chi_p(t)$ in the frequency domain, related to the dielectric function by means of $\epsilon_p(\omega) = 1 + 4\pi\alpha(\omega)$, formally the same as (5.10) with the whole $\chi_p(\omega)$ replaced by the polarization part $\alpha(\omega)$.

Assuming that $\chi_p(\tau)$ in the time domain be a regular function, $\alpha(\omega)$ turns out to be analytic in the $\omega_2 \geq 0$ half-plane, with $\omega = \omega_1 + i\omega_2$. The Cauchy theorem can therefore be applied, selecting the integral contour as in Fig. 5.2: this includes the path along the x -axis with the half-circle avoiding ω_0 , closed in the upper plane $\omega_2 \geq 0$ by the half-circle whose radius is then pushed to infinity. One therefore has:

$$i\pi\alpha(\omega_0) = P \int_{-\infty}^{\infty} d\omega \frac{\alpha(\omega)}{\omega - \omega_0}, \quad (5.19)$$

where P labels the principal value of the integral, as reminded in Appendix 5.8.

Use of (5.10) along with (5.19) provides the self-consistent relation for the polarization dielectric function $\epsilon_p(\omega)$

$$\epsilon_p(\omega) = 1 + \frac{1}{i\pi} P \int_{-\infty}^{\infty} d\omega' \frac{\epsilon_p(\omega') - 1}{\omega' - \omega}, \quad (5.20)$$

that is known as Kramers-Kronig relation. Separation of the real $\epsilon_{p1}(\omega)$ and imaginary $\epsilon_{p2}(\omega)$ parts of the dielectric function yields the explicit expressions:

$$\begin{aligned} \epsilon_{p1}(\omega) &= 1 + \frac{1}{\pi} P \int_{-\infty}^{\infty} d\omega' \frac{\epsilon_{p2}(\omega')}{\omega' - \omega} \\ \epsilon_{p2}(\omega) &= -\frac{1}{\pi} P \int_{-\infty}^{\infty} d\omega' \frac{\epsilon_{p1}(\omega') - 1}{\omega' - \omega}. \end{aligned} \quad (5.21)$$

Equations (5.21) express that real and imaginary parts of the dielectric function are intrinsically related to each other. Considering the physical meaning of the real and imaginary parts of $\epsilon(\omega)$,

Concept

Propagation and absorption behaviors are physically related to each other. Mathematically, by means of the Kramers-Kronig relations.

Notice that identical relations hold for the real $\chi'(\omega)$ and imaginary $\chi''(\omega)$ parts of $\chi(\omega)$, with the formal replacement $\varepsilon_{p1}(\omega) - 1 \rightarrow \chi'(\omega)$ and $\varepsilon_{p2}(\omega) \rightarrow \chi''(\omega)$.

5.2.2.2 Dielectric function and conductivity

As anticipated, medium polarization is not the only manifestation of the system response to an external space- and time-dependent electric field. Indeed, the system may respond as well by means of its own frequency-dependent conductivity to set in a space- and time-dependent current:

$$\mathbf{J}(\mathbf{r}, t) = \int d\mathbf{r}' \int_{-\infty}^t dt' \sigma(\mathbf{r}, \mathbf{r}', t - t') \mathbf{E}(\mathbf{r}', t'), \quad (5.22)$$

This equation connects the current to the total field $\mathbf{E}(\mathbf{r}', t')$. As discussed in Appendix 5.7, the principle of charge conservation leads to the continuity equation $i\mathbf{k} \cdot \mathbf{J}(\mathbf{k}, \omega) = i\omega\rho_c(\mathbf{k}, \omega)$ that relates current $\mathbf{J}(\mathbf{k}, \omega)$ and charge $\rho_c(\mathbf{k}, \omega)$ densities in the (\mathbf{k}, ω) domain of a homogeneous system. The total field $\mathbf{E}(\mathbf{k}, \omega)$ is in turn originated by induced $\rho_{\text{ind}} = -\nabla \cdot \mathbf{P}$ charges as driven by the polarization vector \mathbf{P} , as well as by conduction ρ_c and external ρ_{ext} charges. Thus, the Maxwell equation for $\mathbf{E}(\mathbf{k}, \omega)$ reads

$$i\mathbf{k} \cdot \mathbf{E}(\mathbf{k}, \omega) = 4\pi(-\chi_p(\mathbf{k}\omega)i\mathbf{k} \cdot \mathbf{E}(\mathbf{k}, \omega) + \rho_c(\mathbf{k}, \omega) + \rho_{\text{ext}}(\mathbf{k}, \omega)),$$

or else $i\mathbf{k} \cdot \varepsilon(\mathbf{k}\omega)\mathbf{E}(\mathbf{k}, \omega) = 4\pi\rho_{\text{ext}}(\mathbf{k}, \omega)$, with

$$\varepsilon(\mathbf{k}, \omega) = \varepsilon_p(\mathbf{k}, \omega) - \frac{4\pi}{i\omega} \sigma(\mathbf{k}, \omega). \quad (5.23)$$

In essence, the ω - and k -dependent conductivity $\sigma(\mathbf{k}, \omega)$ is found to contribute to the overall dielectric function along with the polarization dielectric function.

The longitudinal current response can thus be treated in a quite similar manner than the polarization response via the continuity equation, extending all the above results (5.21) to the total dielectric function. In addition, the parity properties of $\chi(\omega)$ imply significant consequences on the real and imaginary parts of $\varepsilon(\omega)$: one finds indeed that $\varepsilon_1(\omega) = \varepsilon_1(-\omega)$ is an even function of frequency whereas $\varepsilon_2(\omega) = -\varepsilon_2(-\omega)$ is odd. As a consequence, the analogue of equation (5.21) for the total dielectric function can be cast in the more expressive form

$$\begin{aligned} \varepsilon_1(\omega) &= 1 + \frac{2}{\pi} P \int_0^\infty d\omega' \frac{\omega' \varepsilon_2(\omega')}{\omega'^2 - \omega^2} \\ \varepsilon_2(\omega) &= -\frac{2\omega}{\pi} P \int_0^\infty d\omega' \frac{\varepsilon_1(\omega') - 1}{\omega'^2 - \omega^2}, \end{aligned} \quad (5.24)$$

after multiplying and dividing by $\omega' + \omega$ and using the parity properties of $\varepsilon(\mathbf{k}, \omega)$.

These are quite general relations between the real and imaginary parts of the frequency-dependent dielectric function, suited for classical and quantum, homogeneous or inhomogeneous, infinite or confined systems.

Quick Questions

Q3. Could $\epsilon_2(\omega)$ be negative at selected frequencies or for given frequency ranges?

Answer - Yes, it can be. In this case, the absorption coefficient turns out to be negative, implying amplification of the electromagnetic field.

Q4. Could $\epsilon_2(\omega)$ be negative at all frequencies?

Answer - No, it cannot be. If it were negative at all frequencies, the electromagnetic radiation would be amplified at all frequencies, leading to energy feeding of the field at indefinitely long times: a situation that is physically meaningless.

Q5. Consider the case with $\epsilon_2(\omega') \rightarrow 0$ for $\omega' \rightarrow \infty$. Tell the expression of $\epsilon_1(\omega)$ for $\omega \rightarrow \infty$.

Answer - By inspection, one immediately sees that $\epsilon_1(\infty) \rightarrow 1$, that is the dielectric function of vacuum.

Q6. Consider the case with $\epsilon_2(\omega) \rightarrow 0$ at all ω values. Express $\epsilon_1(\omega)$.

Answer - By inspection, one immediately sees that $\epsilon_1(\omega) \rightarrow 1$ at all frequencies, that is once again the vacuum dielectric function.

Q7. Consider the case in which $\epsilon_2(\omega)$ is non-vanishing only in the frequency range $\omega_1 < \omega < \omega_2$. Tell the expression of $\epsilon_1(\omega)$ when $\omega < \omega_1$ and $\omega > \omega_2$.

Answer - Since

$$\epsilon_1(\omega) = 1 + \frac{2}{\pi} P \int_{\omega_1}^{\omega_2} d\omega' \frac{\omega' \epsilon_2(\omega')}{\omega'^2 - \omega^2},$$

one sees that $\epsilon_1(\omega) > 1$ while $\omega < \omega_1$, whereas $\epsilon_1(\omega) < 1$ while $\omega > \omega_2$. The physical meaning of this situation is clear after recalling from elementary physics that $\epsilon_1(\omega)$ dictates the phase velocity by means of the relation $\epsilon_1(\omega) = \sqrt{c/v}$, where c is the speed of light in vacuum and v that one inside the material. Therefore, while $\omega > \omega_2$ the phase velocity of light in the medium is larger than c . Of course, this is not in contrast with the principles of the theory of relativity, and might actually occur under suited conditions.

5.2.2.3 Fresnel relations

This section is devoted to recall, from elementary optics, the link among the phenomena of refraction, reflection, and transmission. The link is provided by the Fresnel relations.

Consider a radiation passing from one medium to another. As a reference, consider the case in which one of the two is vacuum. Maxwell equations imply that at the surface separation the tangential components of the electric \mathbf{E} and magnetic \mathbf{H} fields are to satisfy continuity conditions. The same is to occur for the normal components of the electric displacement \mathbf{D} and of magnetic induction \mathbf{B} vectors. Consider the two alternative cases of radiation characterized by an electric field that is either lying onto the interface plane between the two media, or normal to it. The former case refers to s -polarized and the latter to p -polarized light. The case of unpolarized light can be represented as a weighted mix of p - and s -polarized radiation, and the results below can be easily extended. Be i and r the angles that the direction normal to the interface forms with the incident and the refracted light beam, respectively. From elementary optics it is known that the amplitudes r_p and r_s of the p - and s -polarized reflected radiation are related to reflectivity R_p and R_s through $R_p = |r_p|^2$ and $R_s = |r_s|^2$. One finds:

$$\begin{aligned} r_p &= \frac{\tilde{n} \cos r - \cos i}{\tilde{n} \cos r + \cos i} \\ r_s &= \frac{\tilde{n} \cos i - \cos r}{\tilde{n} \cos i + \cos r}, \end{aligned} \quad (5.25)$$

in terms of the complex refraction index $\tilde{n} \equiv n + i\kappa$. In the region where the radiation is fully transmitted, \tilde{n} is a real quantity. Using Snell's law $\sin(i)/\sin(r) = n$, one has:

$$\begin{aligned} r_p &= -\frac{\tan(i-r)}{\tan(i+r)} \\ r_s &= -\frac{\sin(i-r)}{\sin(i+r)}. \end{aligned} \quad (5.26)$$

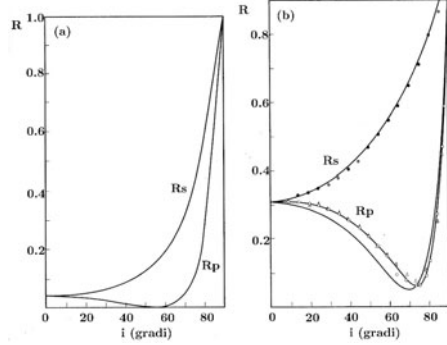
Notice that when $i + r = \pi/2$, thus $r_p = 0$ so that only $r_s \neq 0$. This condition defines the Brewster angle, at which p -polarized light is purely refracted. Under these conditions, one has

$$n = \frac{\sin(i)}{\sin(r)} = \frac{\sin(i)}{\sin(\frac{\pi}{2} - i)} = \tan(i). \quad (5.27)$$

Therefore, a measurement of n as a function of the radiation wavelength can be performed. Equation (5.25) also implies that at normal incidence the reflectivity R and the transmittance T are given by the Fresnel relations

$$R = \frac{(n-1)^2 + \kappa^2}{(n+1)^2 + \kappa^2}$$

Fig. 5.3 Optical properties of dielectric media. Reflectivity R_s of s -polarized light normal to the plane and R_p of p -polarized light parallel to the plane with varying the incidence angle. (a) Prediction for an isotropic transparent material; (b) Data for natural graphite with $n_x = 2.73$, $\kappa_x = 1.4$, $n_z = 1.53$, and $\kappa_z = 0$. Notice that at the Brewster angle the case with $\kappa \neq 0$ corresponds to $R_p \neq 0$ though small. The light wavelength is $\lambda = 6328 \text{ \AA}$. Reprinted with permission from [8]. Copyright (1969) by the American Physical Society



$$T = 1 - R = \frac{4n}{(n+1)^2 + \kappa^2}. \quad (5.28)$$

Fig. 5.3 examines the reflectivity $R_p = |r_p|^2$ and $R_s = |r_s|^2$ as functions of incidence angle. Panel (a) displays the predicted R_p and R_s for an isotropic dielectric medium with $n = 1.5$ and $\kappa = 0$. The predictions are compared with experimental data on natural graphite as reported in panel (b). Determination of n and κ is usually performed by measuring the absorption coefficient in the spectral region where it is significant, and then the reflectivity at normal incidence. In the transparency region, n is usually measured at the Brewster angle and κ from reflectivity measurements. The predictions reported on panel (a) are in qualitative agreement with the data displayed on panel (b), even though it has to be recalled that graphite is anisotropic. In particular, it has $n_x = 2.73$ and $\kappa_x = 1.4$, $n_z = 1.53$ and $\kappa_z = 0$. Notice that the Brewster angle is that for which $R_p = 0$ on panel (a). Comparison with the data in panel (b) shows instead that when $\kappa \neq 0$, R_p at the Brewster angle is finite though small.

5.2.2.4 Static dielectric constant and Lyddane-Sachs-Teller relation

Measurements are often performed in the static $\omega \rightarrow 0$ limit, so that a special attention is conveniently addressed to the static dielectric constant. From (5.24) one has

$$\varepsilon_1(0) = 1 + \frac{2}{\pi} \int_0^\infty d\omega' \frac{\varepsilon_2(\omega')}{\omega'}, \quad (5.29)$$

expressing that

Concept

The static dielectric constant, that is with $\omega \rightarrow 0$, tends to be small as long as the absorption measured by $\epsilon_2(\omega)$ is significant only at high frequencies.

This prediction is confirmed by experimental findings, since the values of the static dielectric constant are indeed very large in metals and progressively smaller in semiconductors and insulators. In fact, the frequencies ω at which significant absorption occurs are dictated by the energy scale of the corresponding excitations: in metals one expects $\hbar\omega \rightarrow (E - \epsilon_f) \rightarrow 0$ since states are immediately available close to the Fermi level, while in semiconductors and insulators one expects that $\hbar\omega > \hbar\omega_0 = E_g$ be larger than the threshold given by the energy gap, that is progressively increasing while passing from semiconductors to insulators.

Consider now the special case in which a material can absorb light only at one single frequency ω_0 , that is in which $\epsilon_2(\omega)$ can be represented by:

$$\epsilon_2(\omega) = C\delta(\omega - \omega_0), \quad (5.30)$$

with C a positive constant. This case corresponds to a simple modeling of a number of systems. Then, (5.29) leads to

$$\epsilon_1(0) = 1 + \frac{2}{\pi} \frac{C}{\omega_0}, \quad (5.31)$$

and using the Kramers-Kronig relations:

$$\begin{aligned} \epsilon_2(\omega) &= \frac{\pi}{2} \omega_0 (\epsilon_1(0) - 1) \delta(\omega - \omega_0) \\ \epsilon_1(\omega) &= 1 + \frac{\omega_0^2 (\epsilon_1(0) - 1)}{\omega_0^2 - \omega^2}. \end{aligned} \quad (5.32)$$

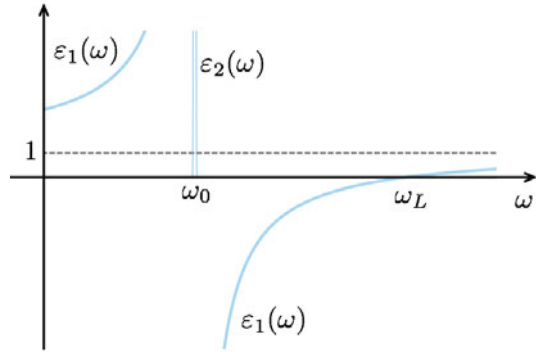
Fig. 5.4 traces the behavior of the imaginary $\epsilon_2(\omega)$ and real $\epsilon_1(\omega)$ parts of the dielectric function within this model. Notice that $\epsilon_1(\omega)$ gets infinitely large values while $\omega \rightarrow \omega_0$. In addition, for $\omega > \omega_0$ $\epsilon_1(\omega)$ switches from negative to positive values as ω becomes larger than a characteristic value ω_L . The latter is evidently defined by imposing the condition $\epsilon_1(\omega_L) = 0$. In particular, from (5.32) one finds that

$$\frac{\omega_L^2}{\omega_0^2} = \epsilon_1(0). \quad (5.33)$$

This is known as the Lyddane-Sachs-Teller relation.

Appendix 5.7 shows that in correspondence of ω_L the Maxwell equations allow a longitudinal solution with vanishing magnetic field. These correspond to collective oscillation modes of the charged particles, as discussed later on. The meaning of the Lyddane-Sachs-Teller relation is clear when the dispersion relation $k^2 = \omega^2 \epsilon(\omega)/c^2$ is recalled. Thus frequencies in the interval (ω_0, ω_L) correspond to imaginary wavevectors and therefore to no propagation. Thus:

Fig. 5.4 Lyddane-Sachs-Teller relation. Dielectric function for a model with one single absorption frequency ω_0 . The real $\epsilon_1(\omega)$ and imaginary $\epsilon_2(\omega)$ parts of the dielectric function are drawn, evidencing the frequency ω_L at which the real part vanishes



Concept

Materials able to absorb light at a finite characteristic frequency ω_0 are completely unavailable to propagation of radiation with frequencies lying in the range (ω_0, ω_L) . The value of ω_L is fixed by the Lyddane-Sachs-Teller relation (5.33), whereas from (5.31) it is seen that larger values of $\epsilon_1(0)$ correspond to smaller values of ω_0 .

Quick Questions

Q8. Might the constant C be negative?

Answer - No, it cannot. If it were so, one would have a system able to amplify the electromagnetic field at the frequency ω_0 and modify the propagation properties at the other frequencies.

Q9. Trace the graphical behavior of $\epsilon_1(\omega)$ for the case analyzed in Q8.

Answer - If $C < 0$ and $1 + 2C/(\pi\omega_0) > 0$ the function would result to be positive smaller than unity at $\omega = 0$. Then it would decrease to vanishing values in the interval $(0, \omega_L)$, vanish at $\omega = \omega_L$, become negative for $\omega > \omega_L$ while tending to infinite values as $\omega \rightarrow \omega_0^-$. At larger values than ω_0 , the function would tend to positive and infinitely large values $\omega \rightarrow \omega_0^+$, asymptotically decreasing to 1 while $\omega \rightarrow \infty$.

5.2.3 Applications and comparison with experimental results

In order to apply the above theoretical findings to real examples, the following procedure can be organized:

Procedure

Step 1. Write the total dielectric function (5.23) $\epsilon(\mathbf{k}, \omega) = \epsilon_p(\mathbf{k}, \omega) + 4\pi i\sigma(\mathbf{k}, \omega)/\omega$, including polarization and conduction terms.

Step 2. Determine the polarization part $\epsilon_p(\omega) = 1 + 4\pi\alpha(\omega)$ from (5.17) through the polarizability, that is the system polarization response to an applied electric field. Use in alternative its Fourier-transformed form $\mathbf{P}(\omega) = \alpha(\omega)\mathbf{E}(\omega)$.

Step 3. Determine the current part $4\pi i\sigma(\mathbf{k}, \omega)/\omega$ from the relation (5.22) for the conductivity, that is the system current response an applied electric field. Use in alternative its Fourier-transformed form $\mathbf{J}(\omega) = \sigma(\omega)\mathbf{E}(\omega)$.

Step 4. Use (5.24) to disentangle and calculate the real and imaginary parts of $\epsilon(\omega)$, that are respectively connected to propagation and absorption.

Step 5. Use (5.15) connecting the complex frequency-dependent dielectric function and refraction index to determine: the absorption coefficient $\eta = 2\omega\kappa/c = \omega\epsilon_2/(nc)$; the wavevector $k_1 = \omega n/c$ providing velocity and dispersion behavior of propagating radiation in the medium.

Step 6. Use the Fresnel relations (5.28) to determine the portion of reflected and transmitted radiation.

Step 7. If applicable, use the Lyddane-Sachs-Teller relation (5.33) to determine the range of radiation frequencies for which the radiation cannot propagate inside the medium, so that the latter is characterized by high reflectivity. In particular, one can determine either one among the following three quantities, once the other two are known: typical absorption frequency ω_0 , frequency ω_L of the propagating longitudinal mode at zero magnetic field, or static dielectric constant $\epsilon_1(0)$.

Examples

The remaining part of the present section is devoted to apply this procedure to understand the optical properties of materials with different behavior with respect to optical or transport phenomena, in particular metals, insulators and polar materials.

Then, Sec. 5.3 extends the theoretical framework to the microscopic treatment of quantum system and Sec. 5.4 to include the effect of interactions of the carriers among themselves and with the lattice vibrations.

5.2.3.1 Dielectric function of metals and experimental findings

The simple model discussed in Sec. 5.2.1 is immediately adapted to describe the optical behavior of metals within the Sommerfeld theory. To this aim, a few modifications are in order. First, the elastic force is negligible and is to be set to zero, or else $\omega_0 = 0$. Then, the mass m and charge e represent the effective mass m^* of the electron at the Fermi level and its charge $-e$. At last, the equation for the classical displacement $y(t)$ is turned into an equation for the change of velocity $\delta\mathbf{v}(t)$ of electrons in the vicinity of the Fermi level. The equation of motion (5.1) thus becomes:

$$m^* \frac{d\delta\mathbf{v}(t)}{dt} + m^* \frac{\delta\mathbf{v}(t)}{\tau} = -e\mathbf{E}(t). \quad (5.34)$$

This can easily be solved in Fourier space as

$$m^* \left(-i\omega + \frac{1}{\tau} \right) \delta\mathbf{v}(\omega) = -e\mathbf{E}(\omega). \quad (5.35)$$

The electric current and the polarization contributions to the dielectric functions are to be determined. As to the current density, this is

$$\mathbf{J}(\omega) = -en\delta\mathbf{v}(\omega) = \frac{e^2 n}{m^* (-i\omega + 1/\tau)} \mathbf{E}(\omega), \quad (5.36)$$

resulting into the electric conductivity $\sigma(\omega)$

$$\sigma(\omega) = \frac{e^2 n}{m^* (-i\omega + 1/\tau)}. \quad (5.37)$$

As to the ionic polarization response to an external electric field, it has already been noticed that this is negligible in metals so that the polarization part of the dielectric function $\epsilon_p(\omega) = 1$ at all frequencies. From (5.23), the total dielectric function is therefore

$$\epsilon(\omega) = 1 - \frac{4\pi n e^2}{m^* \omega (\omega + i/\tau)}. \quad (5.38)$$

Notice that this expression coincides with that obtained from (5.12) after setting $\omega_0 = 0$.

Separation of real and imaginary parts leads to the results

$$\text{Im}(\epsilon(\omega)) = \frac{\omega_p^2}{\omega^2} \frac{\omega\tau}{(\omega\tau)^2 + 1}$$

$$\text{Re}(\varepsilon(\omega)) = 1 - \frac{\omega_p^2}{\omega^2} \frac{(\omega\tau)^2}{(\omega\tau)^2 + 1}, \quad (5.39)$$

where the characteristic frequency ω_p with $\omega_p^2 = 4\pi ne^2/m^*$ has been introduced. This frequency corresponds to a collective motion of electrons, as illustrated by the following simple model. Assume to have a capacitor with free charges between its plates. The simultaneous displacement η of all the charges at the same time towards one of the plates generates on them a surface charge density $\mp ne\eta$. This originates the electric field $E = 4\pi ne\eta$, so that the Newton equation for one electron in between the capacitor is

$$m^* \ddot{\eta} = -4\pi ne^2 \eta. \quad (5.40)$$

This is the equation for a harmonic oscillator for the dynamical variable η . Thus, the electrons perform harmonic oscillations between the plates, at the frequency $\omega_p = \sqrt{4\pi ne^2/m^*}$. This is called plasma frequency. Inserting typical values of the physical constants in metals, one has that $\omega_p \simeq 10^{15} \text{ s}^{-1}$. These values correspond to long wavelengths, of the order of $\lambda \simeq 10^{-5} \text{ cm}$. Thus, the initial assumption that the charges be freely moving within the capacitor is validated by the condition $\omega\tau \gg 1$. In essence,

Concept

Systems of many electrons freely moving under an overall electric field can be characterized by collective, that is long-wavelength or $k \rightarrow 0$, oscillation modes at the plasma frequency ω_p . The corresponding collective excitation is named plasmon.

The theoretical treatment of plasmon collective excitations is beyond the scope of this textbook. However, the implications of the existence of plasmons is discussed hereafter. In fact, the following properties characterize the system as driven by the parameter $\omega\tau$:

Properties

P1. Case with $\omega\tau \gg 1$. The time elapsed between two successive collisions is much longer than the oscillation period of the electromagnetic field: the imaginary part has negligible values and the dielectric function can be approximated by the simple form

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2}. \quad (5.41)$$

The dielectric function is negative for $\omega < \omega_p$ and therefore propagation of the radiation is strongly damped inside the material. In addition, $\varepsilon(\omega) = 0$ at the plasma frequency $\omega = \omega_p$, so that ω_p is the frequency of the longitudinal mode.

P2. **Case with $\omega\tau \ll 1$.** Many collisions occur during one period of oscillation of the electric field: the imaginary part becomes predominant with respect to the real part, and dissipative effects prevail.

In essence,

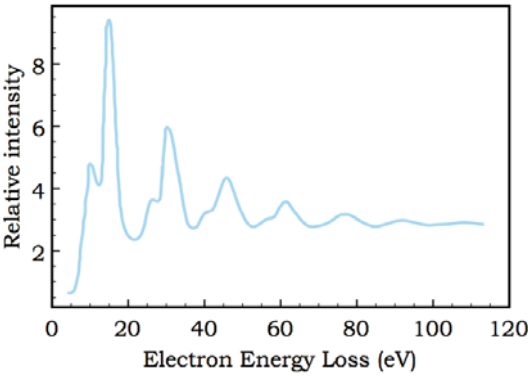
Concept

Metals are characterized by propagation with negligible absorption of electromagnetic radiation with frequencies larger than the plasma frequency of the collective electron modes. Radiation with frequencies $\omega < \omega_p$ below the plasma one, cannot propagate and are instead significantly reflected and damped, unless scattering mechanisms are somehow turned off so that $\tau = 0$.

Plasmons are longitudinal excitations of charged light particles. As discussed in Chap. 2, inelastic scattering is better suited to probe these excitations. Experiments aimed to observe plasmon collective modes are indeed typically performed by bombing a metal layer with protons or α particles. The layer thickness is adjusted to be not too thin to avoid predominance of surface effects and not too thick so that the probe particles can pass over. Indeed, one typically measures the fraction f of particles in the probe beam that undergo the energy loss $\Delta E = E_f - E_i$ from the incoming E_i to the final E_f energy. A typical experimental outcome is displayed in Fig. 5.5 for Al. The relative intensity of the beam that has lost an energy ΔE is peaked at multiple values of $\hbar\omega_p$. The multiple peaks are a manifestation of the quantum nature of this process: plasmons are Bose-like quantum excitations with energy $\hbar\omega_p$, and the energy loss process creates a number of quanta with probabilities, and thus relative intensities, that decrease with increasing the involved energy loss. Notice also that secondary peaks at the left of each main plasmon peak show up. A quantitative analysis shows that the energies of the secondary peaks are a factor $\sqrt{2}$ smaller than that of the primary one: this is a manifestation of plasmon excitations localized at the layer surface, as discussed in problem 5.5.

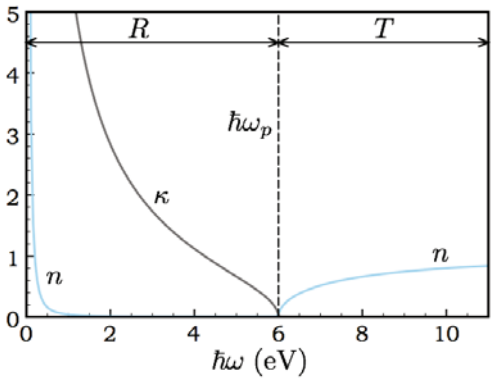
Since the Sommerfeld theory treats the metals as a gas of free electrons subjected to scattering mechanisms, the free-electron gas limit is reached under the condition $\omega\tau \rightarrow \infty$. In this case, expressions (5.39) lead to the determination of n and κ after using the relations $n^2 - \kappa^2 = \epsilon_1(\omega)$ and $2n\kappa = \epsilon_2(\omega)$. Fig. 5.6 displays the absorption coefficient κ and the index of refraction n for such a gas of free electrons. As anticipated, radiation with frequencies $\omega < \omega_p$ below the plasma frequency is typically reflected, whereas the metal is transparent to radiation with frequencies $\omega > \omega_p$ above the plasma frequency. The dependence of the reflectivity on the radiation frequency is indeed illustrated in Fig. 5.7. The theoretical prediction represented in Fig. 5.7 qualitatively agrees with the experimental findings. For comparison, the data referring to Al are displayed in Fig. 5.8, where the value of the

Fig. 5.5 Dielectric function of metals. Plasmon excitations from energy loss experiments in Al. The relative intensity of the beam impinging onto the layer is displayed as a function of the energy loss. The main peaks at multiple values of $\hbar\omega_p$ correspond to plasmon excitations. The secondary peaks on the left of each primary peak are originated by plasmon excitations localized at the layer surface [9]



plasma frequency can be inferred to be about 15 eV/ \hbar . A better qualitative agreement with the data would require the inclusion of interband transitions, which the dip at about 1.4 eV is due to.

Fig. 5.6 Dielectric function of metals. Qualitative sketch of the frequency dependence of the refractive index n and of the absorption coefficient κ for a gas of free electrons. The regions labeled by R and T are characterized by significant reflectivity and transparency, respectively



Quick Questions

Q10. Electromagnetic waves in microwave ovens have frequencies of 2.4 GHz and thus wavelengths ≈ 1.2 mm. Tell which among the following objects you would best avoid to insert in a working microwave oven (and why): a flat metal dish with food on top, a food wrapped in aluminum foil, a tea cup with metallic decorations, a cd-rom [1].

Answer - Typical plasma frequencies in metals are of the order of $\omega_p \approx 5 \cdot 10^{15} \text{ s}^{-1}$, thus $\omega < \omega_p$ for the frequency of microwaves. As a result, microwaves are reflected by usual metals. Therefore, no problems should arise with the flat metal dish: microwaves would enter the food from the top and then be reflected as much as they are from the internal (metallic) oven walls. The food inside the aluminum-foil wrapping won't reasonably cook. Metallic decorations, sharp edges in the tiny holes making the cd-rom, and possible tiny scraps into the aluminum foil should be accurately avoided: microwaves would either drive vigorous high-frequency currents dissipating large powers by Joule effect or, as in the sharp edge case, would accumulate large amounts of charge in tightly shaped spaces, possibly leading to sparks [do not try, please].

Q11. Starting from $\text{Re}(\epsilon(\omega))$ in (5.39), determine ω_L^2 . Is there a limit in which ω_L vanishes?

Answer - Yes, it does. For ω_L one finds

$$\omega_L^2 = \frac{(\omega_p \tau)^2 - 1}{\tau^2},$$

that is real only if $\omega_p \tau > 1$. When $\omega_p \tau = 1$, $\omega_L = 0$. Notice that for $\omega_p \tau < 1$ instead, ω_L is imaginary.

Q12. Using the results of Q11, discuss the dependence of $\epsilon_2(\omega_L)$ on $\omega_p \tau$. Is there a limit of strongly enhanced absorption?

Answer - Inserting the expression of ω_L obtained in Q11 into (5.39), one finds that

$$\epsilon_2(\omega_L) = \frac{(\omega_p \tau)^2}{(\omega_p \tau)^2 - 1} \frac{\sqrt{(\omega_p \tau)^2 - 1}}{(\omega_p \tau)^2} = \frac{1}{\sqrt{(\omega_p \tau)^2 - 1}}.$$

One then sees that $\epsilon_2(\omega_L)$ gets infinitely large values as $\omega_p \tau \rightarrow 1$ and decreases with increasing $\omega_p \tau$.

5.2.3.2 Dielectric function of insulators and experimental findings

The full simple model discussed in Sec. 5.2.1 is straightforward to describe the optical behavior of insulators. This time the charged particles capable of motion are the ions, so that M represents the ion mass and Q their charge. Equations (5.13) provide the real and imaginary parts of the dielectric function. The whole reasoning about the plasma frequency of electrons can be transferred to ions, so that now $\omega_p^2 = 4\pi N Q^2 / M$ is defined, with N the number of ions per unit volume. Notice that the ratio $\omega_p / \omega_p = \sqrt{N Q^2 m^* / (n e^2 M)} \ll 10^{-2}$ of the ions to the electrons plasma

frequency is very small, considering that indeed $N \simeq n$ and $|Q| \simeq e$. Following the outlined procedure, all the different quantities can be calculated. Fig. 5.9 displays for example the frequency dependence of n and κ . Notice that the imaginary part is significant in the vicinity of ω_0 while it becomes rapidly negligible at all the other frequencies. In fact, the average width of the $\kappa(\omega)$ curve is $2/\tau$. If collisions are rare, that is in the $\omega\tau \rightarrow \infty$ limit, the frequency range where κ is significant, narrows around ω_0 . In this case, the real part becomes negative in the interval (ω_0, ω_L) , with ω_L being determined from the condition $\epsilon_1(\omega_L) = 0$. Thus:

$$\omega_L = \sqrt{\omega_0^2 + \omega_p^2}, \quad (5.42)$$

Radiation with frequency in the range (ω_0, ω_L) cannot propagate inside the insulator and is thus mainly reflected. This is the region III in Fig. 5.9.

Problem 5.7 elaborates the calculation of the values ω_{\pm} at which $\epsilon_1(\omega_{\pm}) = 0$ in the more general case: it is indeed found that the imaginary part is not generally negligible in the vicinity of ω_{\pm} .

Fig. 5.7 Dielectric function of metals. Qualitative sketch of the frequency dependence of reflectivity for a gas of free electrons. The reflectivity drastically drops down above the plasma frequency ω_p

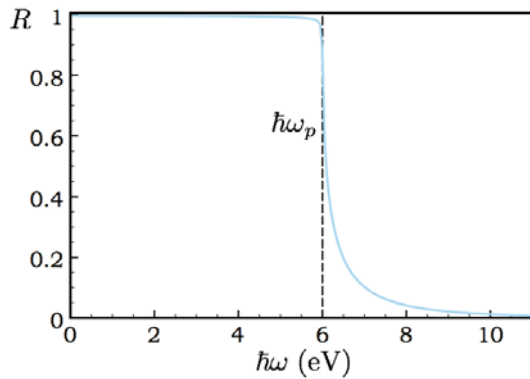


Fig. 5.8 Dielectric function of metals. Experimental data for the frequency dependence of reflectivity of Al. The predictions in Fig. 5.7 are in qualitative agreement with the data. The dip at about 1.4 eV is due to an interband transition. The drastic drop of reflectivity at about 15 eV infers the value of plasma frequency in this material [10]

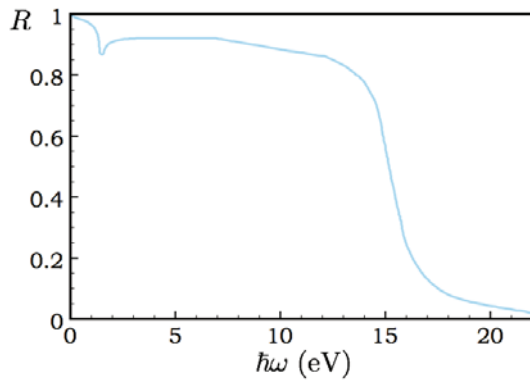
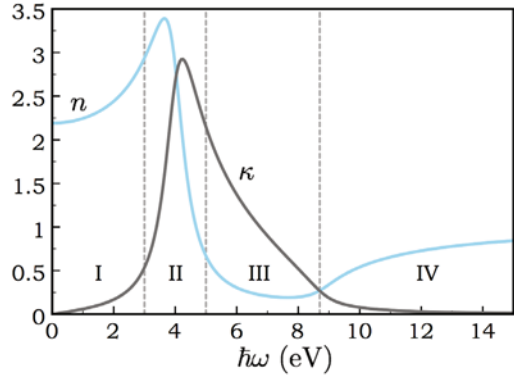


Fig. 5.9 Dielectric function of insulators. Frequency dependence of the real n and imaginary κ parts of the complex refractive index. The material is highly transparent in regions I and IV, strong absorber in region II, and highly reflecting in region III. In this case $\hbar\omega_0 = 4$ eV and $\hbar\tau^{-1} = 1$ eV. From lower to higher energies, the edge of the different regions are at $\hbar\omega_0 \pm \hbar\tau^{-1}$ and the third at $\hbar\omega_L$, and marked by the vertical dashed lines



5.2.3.3 Dielectric function of polar materials and experimental findings

In the case of polar materials, the moving particles are at least two, that are the ions in the elementary cell. Therefore, the simple model discussed in Sec. 5.2.1 needs to be extended to include this essential detail. The simplest model extension is to consider a polar material with two atoms in the elementary cell and in the presence of an external time-dependent electric field $\mathbf{E}(t)$. The two ions have opposite charge, so that their equations of motion are

$$\begin{aligned} M_+ \frac{d^2 \mathbf{u}_+}{dt^2} &= -k(\mathbf{u}_+ - \mathbf{u}_-) + e\mathbf{E}(t) \\ M_- \frac{d^2 \mathbf{u}_-}{dt^2} &= k(\mathbf{u}_+ - \mathbf{u}_-) - e\mathbf{E}(t), \end{aligned} \quad (5.43)$$

with \mathbf{u}_+ and \mathbf{u}_- the out-of-equilibrium displacements of the ions with masses M_+ and M_- . The relative displacement vector $\mathbf{w} = \mathbf{u}_+ - \mathbf{u}_-$ and center of mass vector $\mathbf{R}_{CM} = (M_+ \mathbf{u}_+ + M_- \mathbf{u}_-)/(M_+ + M_-)$ are conveniently introduced, along with the reduced mass M according to $M^{-1} = M_+^{-1} + M_-^{-1}$. The two expressions (5.43) can therefore be cast in more convenient form. Summing them up one obtains $d^2 \mathbf{R}_{CM}(t)/(dt^2) = 0$, expressing the fact that the center of mass of the two ions moves freely: and in particular it remains at rest if so it was at the beginning of the motion. Subtracting them up, one obtains for the relative displacement vector:

$$\frac{d^2 \mathbf{w}}{dt^2} = -\frac{k}{M} \mathbf{w} + \frac{e}{M} \mathbf{E}. \quad (5.44)$$

This equation of motion has been obtained within a specific case. In fact, it can be extended to more general situations after writing:

$$\frac{d^2 \mathbf{w}(t)}{dt^2} = b_{11} \mathbf{w}(t) + b_{12} \mathbf{E}(t) \quad (5.45)$$

Table 5.1 Polar materials. Values of the static ϵ_0 and high-frequency ϵ_∞ dielectric constant, along with the transverse ω_T and longitudinal ω_L phonon frequencies, for the selected polar samples at room temperature listed in the first column. The values of ω_T are from experiments and those of ω_L derived from the Lyddane-Sachs-Teller relation. Data are referred from textbook [3]. For comparison, the data for C, Si, and Ge are also reported, for which no splitting manifests between ω_T and ω_L , so that $\epsilon_0 = \epsilon_\infty$

| Material | ϵ_0 | ϵ_∞ | $ \omega_T (\times 10^{13} \text{ s}^{-1}) $ | $ \omega_L (\times 10^{13} \text{ s}^{-1}) $ |
|----------|--------------|-------------------|--|--|
| LiH | 12.9 | 3.6 | 11.0 | 21.0 |
| LiF | 8.9 | 1.9 | 5.8 | 12.0 |
| LiCl | 12.0 | 2.7 | 3.6 | 7.5 |
| LiBr | 13.2 | 3.2 | 3.0 | 6.1 |
| NaF | 5.1 | 1.7 | 4.5 | 7.8 |
| NaCl | 5.9 | 2.25 | 3.1 | 5.0 |
| NaBr | 6.4 | 2.6 | 2.5 | 3.9 |
| KF | 5.5 | 1.5 | 3.6 | 6.1 |
| KCl | 4.85 | 2.1 | 2.7 | 4.0 |
| KI | 5.1 | 2.7 | 1.9 | 2.6 |
| AgCl | 12.3 | 4.0 | 1.9 | 3.4 |
| AgBr | 13.1 | 4.6 | 1.5 | 2.5 |
| MgO | 9.8 | 2.95 | 7.5 | 14.0 |
| GaP | 10.7 | 8.5 | 6.9 | 7.6 |
| GaAs | 12.9 | 10.9 | 5.1 | 5.5 |
| GaSb | 16.1 | 14.4 | 4.3 | 4.6 |
| InP | 12.4 | 9.6 | 5.7 | 6.5 |
| InAs | 14.9 | 12.3 | 4.1 | 4.5 |
| InSb | 17.7 | 15.6 | 3.5 | 3.7 |
| C | 5.5 | 5.5 | 25.1 | 25.1 |
| Si | 11.7 | 11.7 | 9.9 | 9.9 |
| Ge | 15.8 | 15.8 | 5.7 | 5.7 |

where the quantities b_{11} and b_{12} are to be determined, depending on the specific material. Equation (5.45) has the same form as (5.44), but it is more general. In addition, it bears the same structure as (5.1), but a missing viscous term. In similar manner, the expression for the polarization is conveniently extended to explicitly include the ionic polarization, that is

$$\mathbf{P}(t) = b_{12}\mathbf{w}(t) + b_{22}\mathbf{E}(t).$$

(5.46)

where $b_{12}\mathbf{w}(t)$ accounts for the ions displacement and $b_{22}\mathbf{E}(t)$ for their polarizability. Notice also that no forces due to ions in neighboring cells appear in (5.43): this result reflects the findings of Chap. 3, where the optical phonons in the $k \rightarrow 0$ limit have been seen to represent displacements of the ions in the cell, their center of mass remaining at rest.

Equations (5.45) and (5.46) represent a macroscopic model to calculate the dielectric function of ionic crystals. The expressions of the constants b_{ij} are to be

identified though. To this aim, consider the case in which $\mathbf{w} \cdot \mathbf{E} = 0$: (5.45) leads to $b_{11} = -\omega_T^2$, with ω_T the frequency of the transverse optical phonon, typically lying in the infrared region. The remaining constants are determined as follows. The Fourier transform of (5.45) leads to

$$\mathbf{w}(\omega) = -\frac{b_{12}}{\omega^2 - \omega_T^2} \mathbf{E}(\omega),$$

that can be substituted into the Fourier transform of (5.46) to yield:

$$\mathbf{P}(\omega) = \left(b_{22} - \frac{b_{12}^2}{\omega^2 - \omega_T^2} \right) \mathbf{E}(\omega). \quad (5.47)$$

This relation provides the dielectric function of the material, that is

$$\varepsilon(\omega) = 1 + 4\pi \left(b_{22} - \frac{b_{12}^2}{\omega^2 - \omega_T^2} \right). \quad (5.48)$$

The coefficients b_{22} and b_{12} result in terms of optical constants of the material as follows. One has that $\varepsilon(0) = \varepsilon_0$ is the static dielectric constant. In addition, one has that $\varepsilon(\omega \rightarrow \infty) = 1 + 4\pi b_{22} \equiv \varepsilon_\infty$ is the dielectric constant at frequencies much larger than the transverse optical phonon frequency ω_T ($\omega \gg \omega_T$), and much smaller than the frequencies ω_{el} typical of electronic modes ($\omega \ll \omega_{el}$). Therefore, one has $(\varepsilon_0 - \varepsilon_\infty)\omega_T^2 = b_{12}^2$. In conclusion, the dielectric function can be expressed as

$$\varepsilon(\omega) = \varepsilon_\infty - \frac{(\varepsilon_0 - \varepsilon_\infty)\omega_T^2}{\omega^2 - \omega_T^2}. \quad (5.49)$$

This relation is similar to (5.12), but the viscous term. For this reason the dielectric function has only a real part. The dielectric function $\varepsilon(\omega)$ gets infinitely large values while the frequency approaches ω_T , and negative values in the interval (ω_T, ω_L) . Using the definition $\varepsilon(\omega_L) = 0$, one finds that $\omega_L = \omega_T \sqrt{\varepsilon_0/\varepsilon_\infty}$. Therefore, the Liddane-Sachs-Teller for ionic crystals can be cast in the form

$$\frac{\omega_L^2}{\omega_T^2} = \frac{\varepsilon_0}{\varepsilon_\infty}. \quad (5.50)$$

Finally, the propagation of radiation is dictated by the dispersion relation $\omega(k)$, that is provided by

$$k^2 = \frac{\omega^2}{c^2} \varepsilon(\omega). \quad (5.51)$$

In the interval (ω_T, ω_L) the dielectric function is negative and therefore no propagation of radiation may occur inside the material. The spectrum is composed of two parts. In the interval $(0, \omega_T)$ has a linear behavior $\omega(k) \simeq ck/\sqrt{\varepsilon_0}$ in the long wavelength limit $k \rightarrow 0$ and $\omega(k) \simeq \omega_T$ in the $k \rightarrow \infty$ limit. In the interval (ω_L, ∞) , the linear constant behavior $\omega(k) \simeq \omega_L$ is observed in the $k \rightarrow 0$ limit and the linear be-

havior $\omega(k) = ck/\sqrt{\epsilon_\infty}$ in the $k \rightarrow \infty$ limit. In between ω_T and ω_L , no radiation can propagate, as already discussed. Fig. 5.10 sketches this behavior, evidencing that

Concept

Two branches characterize the dispersion relation of light in polar materials: longitudinal and transverse modes are coupled at intermediate range of k and decouple in the $k \rightarrow 0$ and $k \rightarrow \infty$ limits: here, they evolve towards the bare longitudinal and transverse optical modes. This behavior is a general manifestation occurring whenever two discrete levels exists that are coupled, leading to crossing and anticrossing. In between ω_T and ω_L , a polar material is characterized by full reflectivity.

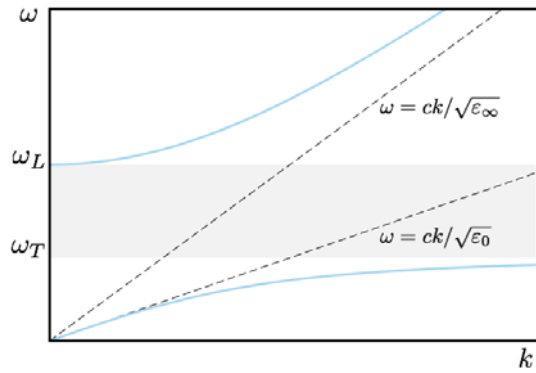


Fig. 5.10 Dielectric function of polar materials. The dispersion relation $\omega(k)$ showing the coupling of the longitudinal and transverse optical modes and the region of total reflectivity (see text)

Quick Questions

Q13. Consider the equation of motion (5.45) and add a viscous term. How is the dielectric function modified?

Answer - One has:

$$\epsilon(\omega) = \epsilon_\infty - \frac{(\epsilon_0 - \epsilon_\infty)\omega_T^2}{\omega^2 - \omega_T^2 + i\omega/\tau},$$

getting i.e. the same structure as (5.12).

Q14. Mirrors exist for makeups, that allow to choose between incandescent and fluorescent light. Say why I should makeup myself in the same light which I suppose to be exposed to during the daytime [1].

Answer - Incandescent light contains more red and less blue light, while fluorescent and sun lights contain more blue and violet light. If I makeup under incandescent light, I could underestimate how much green or blue I'm putting on my face, and viceversa with sunlight I could underestimate how much red.

Q15. White cotton fabrics tend with age to absorb more blue light and begin to look yellow. While bleaching does not solve the problem, modern laundry detergents exist which brighten the cotton fabric as if it were brand new. Guess how this work [1].

Answer - In order to make the cotton white again, it must emit more blue light. One easy solution is to blueing the fabric. Notice also that more recent solutions use fluorescent dyes, designed to absorb into the ultraviolet part of the spectrum (at higher frequency, indeed) and re-emit in the blue part of the spectrum (at lower frequency, as it must generally occur with fluorescence).

5.3 Quantum models for the dielectric function

The formulation of the dielectric function theory in a quantum framework conceptually follows the phenomenological treatment adopted so far. The response of the system to the electromagnetic external field requires quantum tools, and one of those is perturbation theory. Indeed, the lowest order in perturbation theory is hereafter used to understand the way the theory works for weakly interacting particles treated as single particles in the self-consistent field discussed in Chap. 2. Appearing as a special case, in Chap. 6 Density Function Theory is in fact seen to elegantly map a strongly interacting many-particles system into a weakly interacting one solved by fictitious single-particle orbitals. Thus, the results of the present section are recovered and extended in Chap. 6 to the treatment of strongly interacting systems.

Consider to reduce the many particle problem to that of a set of single particles in the self-consistent field. The strategy here is to perform a microscopic calculation of the quantum-mechanical probability that transitions may occur between pairs of states while the radiation propagates inside the material, by using the perturbation theory recalled in Appendix 3.18 of Chap. 3. Then, the result is compared with the macroscopic result elaborated so far for the dissipated power. By inspection, the imaginary part of the dielectric function is obtained and, by Kramers-Kronig relation, the real part.

Appendix 5.7 demonstrates that a dispersive system in frequency and wavenumber is characterized by the existence of two types of solutions of Maxwell equations. These are transverse and longitudinal solutions. The former are defined by the condition that wavenumber \mathbf{q} , electric field $\mathbf{E}(\mathbf{q}, \omega)$, and magnetic field $\mathbf{B}(\mathbf{q}, \omega)$ are all perpendicular to each other, the mode having the dispersion relation $q^2 = \omega^2 / (c^2 \epsilon(\mathbf{q}, \omega))$. The latter are defined by the condition $\mathbf{q} \parallel \mathbf{E}(\mathbf{q}, \omega)$

and $\mathbf{B}(\mathbf{q}, \omega) = 0$, with dispersion relation given by $\varepsilon(\mathbf{q}, \omega) = 0$. Besides, the total electric field $\mathbf{E}(\mathbf{q}, \omega)$ is related to the external field $\mathbf{E}_{\text{ext}}(\mathbf{q}, \omega)$ by the relation $\mathbf{E}(\mathbf{q}, \omega) = \mathbf{E}_{\text{ext}}(\mathbf{q}, \omega)/\varepsilon(\mathbf{q}, \omega)$. It is also known that the Maxwell equations in the quasi-stationary limit reduce to longitudinal equations. Thus, the response of a system to an electromagnetic field can be either transverse or longitudinal. Accordingly, a transverse $\varepsilon_T(\mathbf{q}, \omega)$ and a longitudinal $\varepsilon_L(\mathbf{q}, \omega)$ dielectric functions can be defined to calculate the system response, the two coinciding in the long-wavelength limit, as discussed later on in this section.

The longitudinal and transverse dielectric functions are therefore introduced in the following. The electronic states of the unperturbed Hamiltonian are single-particle states obtained as solution of the eigenvalues equation

$$\begin{aligned} H_0 &= \frac{p^2}{2m} + V(\mathbf{r}) \\ H_0 \psi_j &= E_j \psi_j, \end{aligned} \quad (5.52)$$

with $V(\mathbf{r})$ the self-consistent single-particle potential.

5.3.1 Longitudinal dielectric function

Consider the external longitudinal field $U_{\text{ext}}(\mathbf{r}, t) = U_{\text{ext}}^0(\mathbf{q}, \omega)e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)} + c.c.$. By polarization and/or conduction a total field is generated in the medium, that is $U_{\text{tot}}(\mathbf{r}, t) = U_{\text{tot}}^0(\mathbf{q}, \omega)e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)} + c.c.$, so that the Hamiltonian can be cast into the form

$$H = H_0 + U_{\text{tot}}(\mathbf{r}, t), \quad (5.53)$$

where H_0 is defined as in (5.52). The external electric potential is $\phi_{\text{ext}}(\mathbf{r}, t) = -U_{\text{ext}}(\mathbf{r}, t)/e$ and the total one $\phi_{\text{tot}}(\mathbf{r}, t) = -U_{\text{tot}}(\mathbf{r}, t)/e$. The corresponding electric fields are $\mathbf{E}_{\text{ext}} = -\nabla\phi_{\text{ext}}(\mathbf{r}, t) = \mathbf{E}_{\text{ext}}^{(0)}e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)} + c.c.$ and $\mathbf{E}_{\text{tot}} = -\nabla\phi_{\text{tot}}(\mathbf{r}, t) = \mathbf{E}_0e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)} + c.c.$ with $E_0 = iqU_{\text{tot}}^0(\mathbf{q}, \omega)/e$. Appendix 5.7 recalls that for longitudinal fields one has

$$\begin{aligned} \mathbf{E}_0 &= \frac{\mathbf{E}_{\text{ext}}}{\varepsilon(\mathbf{q}, \omega)} \\ U_{\text{tot}}^0(\mathbf{q}, \omega) &= \frac{U_{\text{ext}}^0(\mathbf{q}, \omega)}{\varepsilon(\mathbf{q}, \omega)}. \end{aligned} \quad (5.54)$$

Following the usual procedure set within the time-dependent perturbation theory, the probability for unit time can be calculated that a transition may occur from state ψ_i to state ψ_j or viceversa from state ψ_j to state ψ_i , by absorption or emission of only one energy quantum $\hbar\omega$ of the external perturbation, respectively. The two probabilities $P_{i \rightarrow j}$ and $P_{j \rightarrow i}$ are

$$\begin{aligned}
P_{i \rightarrow j} &= \frac{2\pi}{\hbar} |\langle \psi_j | U_{tot}^0 e^{i\mathbf{q} \cdot \mathbf{r}} | \psi_i \rangle|^2 \delta(E_j - E_i - \hbar\omega) \\
P_{j \rightarrow i} &= \frac{2\pi}{\hbar} |\langle \psi_i | U_{tot}^0 e^{-i\mathbf{q} \cdot \mathbf{r}} | \psi_j \rangle|^2 \delta(E_i - E_j + \hbar\omega).
\end{aligned} \tag{5.55}$$

Many electrons are present in the system, so that the overall probability for transitions between all possible pairs of states is to be accounted for. This task is accomplished by multiplying $P_{i \rightarrow j}$ times the Fermi quantum distribution $f(E_i)(1 - f(E_j))$, representing the joined probability that state i be occupied and state j be empty. In similar manner, $P_{j \rightarrow i}$ is multiplied by $f(E_j)(1 - f(E_i))$. The two probabilities with the corresponding multiplication factors are then to be subtracted, to have the net balance between absorption and emission. In any event, after summing over all the possible states ψ_i and ψ_j , one finally obtains:

$$\begin{aligned}
W(\mathbf{q}, \omega) &= 2 \frac{2\pi}{\hbar} \sum_{i,j} |U_{tot}^0 \langle \psi_j | e^{i\mathbf{q} \cdot \mathbf{r}} | \psi_i \rangle|^2 \\
&\times \delta(E_j - E_i - \hbar\omega) (f(E_i) - f(E_j)).
\end{aligned} \tag{5.56}$$

This is the total transition probability for absorption and emission processes between all the pairs of states, that are separated by the energy difference $|E_j - E_i| = \hbar\omega$ corresponding to the energy of one single quantum exchanged with the external field. The factor of 2 accounts for the spin degrees of freedom. Each transition requires an energy $\hbar\omega$ and thus the dissipated power is

$$P(\mathbf{q}, \omega) = \hbar\omega W(\mathbf{q}, \omega). \tag{5.57}$$

The simple macroscopic model in Sec. 5.2.1 has already revealed that the dissipated power is related to the imaginary part of the response function. In the microscopic theory, the dissipated power is $\int_V \mathbf{J}(\mathbf{r}, t) \cdot \mathbf{E}(\mathbf{r}, t) d\mathbf{r}$. Now, in the gauge under consideration

$$\mathbf{J}(\mathbf{r}, t) = \sigma(\mathbf{q}, \omega) E_0 \hat{\mathbf{q}} e^{[i(\mathbf{q} \cdot \mathbf{r} - \omega t)]} + \text{c.c.}, \tag{5.58}$$

with $\hat{\mathbf{q}}$ the unit vector identifying the direction \mathbf{q} . Therefore, the power is

$$\int_V \mathbf{J}(\mathbf{r}, t) \cdot \mathbf{E}(\mathbf{r}, t) d\mathbf{r} = 2\sigma_1(\mathbf{q}, \omega) |E_0|^2 V, \tag{5.59}$$

with $\sigma_1(\mathbf{q}, \omega)$ the real part of $\sigma(\mathbf{q}, \omega)$. By comparison of (5.59) with (5.57), one obtains the expression

$$\sigma_1(\mathbf{q}, \omega) = \frac{e^2}{2V} \frac{\hbar\omega W(\mathbf{q}, \omega)}{q^2 |U_{tot}^0|^2}. \tag{5.60}$$

By inspection of (5.23), the imaginary part of the dielectric function turns out to be:

$$\varepsilon_{2,L}(\mathbf{q}, \omega) = \frac{8e^2\pi^2}{q^2 V} \sum_{i,j} |\langle \psi_j | e^{i\mathbf{q} \cdot \mathbf{r}} | \psi_i \rangle|^2 (f(E_i) - f(E_j) \delta(E_j - E_i - \hbar\omega)). \tag{5.61}$$

The real part can thus be obtained by means of Kramers-Kronig relations, that is

$$\varepsilon_{1,L}(\mathbf{q}, \omega) = 1 + \frac{8\pi e^2}{q^2 V} \sum_{i,j} |\langle \psi_j | e^{i\mathbf{q}\cdot\mathbf{r}} | \psi_i \rangle|^2 \frac{f(E_i) - f(E_j)}{E_j - E_i - \hbar\omega}. \quad (5.62)$$

The two relations (5.61) and (5.62) can be assembled into

$$\varepsilon_L(\mathbf{q}, \omega) = 1 + \frac{8\pi e^2}{q^2 V} \sum_{i,j} |\langle \psi_j | e^{i\mathbf{q}\cdot\mathbf{r}} | \psi_i \rangle|^2 \frac{f(E_i) - f(E_j)}{E_j - E_i - \hbar\omega - i\eta}, \quad (5.63)$$

where it is intended that $\eta \rightarrow 0^+$ along with the mathematical result

$$\frac{1}{x + i\eta} = P \frac{1}{x} + i\pi \delta(x). \quad (5.64)$$

The power dissipated is the function that is measured in all the energy loss-like experiments illustrated in the previous Chapters. The expression of $P(\mathbf{q}, \omega)$ can also be cast in the form

$$\begin{aligned} P(\mathbf{q}, \omega) &= V \frac{\omega q^2}{2\pi e^2} \varepsilon_{2,L}(\mathbf{q}, \omega) |U_{tot}^0(\mathbf{q}, \omega)|^2 \\ &= V \frac{\omega q^2}{2\pi e^2} \frac{\varepsilon_{2,L}(\mathbf{q}, \omega)}{|\varepsilon_L(\mathbf{q}, \omega)|^2} |U_{ext}^0(\mathbf{q}, \omega)|^2 \\ &= V \frac{\omega q^2}{2\pi e^2} \left[-\text{Im} \left(\frac{1}{\varepsilon_L(\mathbf{q}, \omega)} \right) \right] |U_{ext}^0(\mathbf{q}, \omega)|^2, \end{aligned} \quad (5.65)$$

evidencing its connection with the imaginary part of the inverse longitudinal dielectric function: the power dissipated is maximum when $-\text{Im}(1/\varepsilon_L(\mathbf{q}, \omega))$ does.

A best matching with the classical expression (5.12) from the simple model can be obtained as follows. Pick up a pair of states i and j . Their contribution X_{ij} to $\varepsilon_L(\mathbf{q}, \omega)$ can be cast in the form

$$\begin{aligned} X_{ij} &= \frac{8\pi e^2}{q^2 V} |\langle \psi_j | e^{i\mathbf{q}\cdot\mathbf{r}} | \psi_i \rangle|^2 \frac{2(E_j - E_i)(f(E_i) - f(E_j))}{(E_j - E_i)^2 - (\hbar\omega + i\eta)^2} \\ &= \frac{16\pi e^2}{q^2 V \hbar} |\langle \psi_j | e^{i\mathbf{q}\cdot\mathbf{r}} | \psi_i \rangle|^2 \frac{(\omega_j - \omega_i)(f(E_i) - f(E_j))}{(\omega_j - \omega_i)^2 - \omega^2 - 2i\omega\eta\hbar^{-1}}. \end{aligned} \quad (5.66)$$

The numerator of (5.66) is positive because $(E_j - E_i)(f(E_i) - f(E_j))$ does. For just dimensional reasons, the latter can be written as $\bar{\omega}_{ij}^2(\mathbf{q})$. Setting $\tau^{-1} = 2\eta\hbar^{-1}$, one finds that X_{ij} has the same form as in (5.12). The quantum version of the longitudinal dielectric function can therefore be written as

$$\varepsilon_L(\mathbf{q}, \omega) = 1 + \sum_{i,j} \frac{\bar{\omega}_{ij}^2(\mathbf{q})}{\omega_{ij}^2 - \omega^2 - i\omega/\tau}. \quad (5.67)$$

The correspondence with the classical form is completed by identifying the natural frequency ω_0 with the transition frequency ω_{ij} and the plasma frequency with $\bar{\omega}_{ij}(\mathbf{q})$. The form (5.67) enlightens what has been anticipated on physical arguments, that is

Concept

Weakly excited systems in matter can be represented as a set of independent damped harmonic oscillators: their frequency ω_{ij} is related to the energy needed to excite the state and the relaxation time τ describes the state lifetime as hampered by the interactions of the state with the surroundings.

5.3.2 Transverse dielectric function

Treating the radiation in quasi-classical approximation, the starting Hamiltonian is

$$H = \frac{1}{2m} \left(\mathbf{p} + \frac{e}{c} \mathbf{A}(\mathbf{r}, t) \right)^2 + V(\mathbf{r}), \quad (5.68)$$

where $\mathbf{A}(\mathbf{r}, t)$ is the vector potential associated to the radiation and satisfying the condition $\nabla \cdot \mathbf{A}(\mathbf{r}, t) = 0$ in the Coulomb gauge. To first order in perturbation theory the Hamiltonian is

$$H = H_0 + \frac{e}{mc} \mathbf{A}(\mathbf{r}, t) \cdot \mathbf{p}. \quad (5.69)$$

Be $\mathbf{A}(\mathbf{r}, t) = A_0 \mathbf{e} \left\{ e^{i(\mathbf{q} \cdot \mathbf{r} - \omega t)} + c.c. \right\}$ with \mathbf{e} the unit vector in direction of the polarization and \mathbf{q} the radiation wavenumber with $\mathbf{e} \cdot \mathbf{q} = 0$, and A_0 the amplitude of the vector potential. Therefore, the electric field is given by $\mathbf{E}(\mathbf{r}, t) = -c^{-1} \partial \mathbf{A}(\mathbf{r}, t) / \partial t$. Following the same procedure as in the calculation of the imaginary parte of the longitudinal dielectric function, one finds

$$\begin{aligned} \varepsilon_{2,T}(\mathbf{q}, \omega) &= \frac{8\pi e^2}{m^2 \omega^2} \frac{1}{V} \sum_{i,j} |\langle \psi_j | e^{i\mathbf{q} \cdot \mathbf{r}} \mathbf{e} \cdot \mathbf{p} | \psi_i \rangle|^2 \\ &\times \delta[E_j - E_i - \hbar \omega] (f(E_i) - f(E_j)). \end{aligned} \quad (5.70)$$

The Kramers-Kronig relation then yields $\varepsilon_{1,T}(\mathbf{q}, \omega)$, that is

$$\varepsilon_{1,T}(\mathbf{q}, \omega) = 1 + \frac{8\pi e^2}{m^2} \frac{1}{V} \sum_{i,j} \frac{|\langle \psi_j | e^{i\mathbf{q} \cdot \mathbf{r}} \mathbf{e} \cdot \mathbf{p} | \psi_i \rangle|^2}{(E_j - E_i)^2 / \hbar^2} \frac{f(E_i) - f(E_j)}{E_j - E_i - \hbar \omega}. \quad (5.71)$$

The two expressions (5.70) and (5.71) can be assembled into the following

$$\varepsilon_T(\mathbf{q}, \omega) = 1 + \frac{8\pi e^2}{m^2} \frac{1}{V} \sum_{i,j} \frac{|\langle \psi_j | e^{i\mathbf{q} \cdot \mathbf{r}} \mathbf{e} \cdot \mathbf{p} | \psi_i \rangle|^2}{(E_j - E_i)^2 / \hbar^2} \frac{f(E_i) - f(E_j)}{E_j - \varepsilon_i - \hbar \omega - i\eta}. \quad (5.72)$$

Notice that the $\mathbf{q} \rightarrow 0$ limiting behavior of (5.72) can be easily found by expanding $e^{i\mathbf{q}\cdot\mathbf{r}}$ in power series inside (5.63) and using the Heisenberg equation

$$i\hbar \frac{\mathbf{p}}{m} = [\mathbf{r}, H_0].$$

Quick Questions

Q16. Consider $\epsilon_L(q, \omega)$ as in (5.63), and set $\omega \neq 0$. Is there a non-vanishing contribution when the states i and j are coincident?

Answer - No. In (5.63) one immediately sees that $f(E_i) = f(E_j)$ and therefore the contribution of such pairs of states to $\epsilon_L(q, \omega)$ is zero.

Q17. Write the dielectric function $\epsilon_L(q, \omega)$ at $T = 0$, for a two-level system with energies E_1 and E_2 and, say, $E_1 < E_2$. Consider the state E_1 as occupied.

Answer - Insertion of $f(E_1) = 1$ and $f(E_2) = 0$ in (5.63) yields

$$\epsilon_L(\mathbf{q}, \omega) = 1 + \frac{16\pi e^2}{q^2} |\langle \psi_1 | e^{i\mathbf{q}\cdot\mathbf{r}} | \psi_2 \rangle|^2 \frac{\Delta E}{(\Delta E)^2 - (\hbar\omega + i\eta)^2},$$

where $\Delta E = E_2 - E_1 = \hbar\omega_0$. Simple algebraic manipulations lead to

$$\begin{aligned} \epsilon_L(\mathbf{q}, \omega) &= 1 + A(\mathbf{q}) \frac{1}{\omega_0^2 - \omega^2 - 2i\eta\omega/\hbar} \\ A(\mathbf{q}) &= \frac{16\pi e^2}{q^2 \hbar^2} \Delta E |\langle \psi_1 | e^{i\mathbf{q}\cdot\mathbf{r}} | \psi_2 \rangle|^2. \end{aligned} \quad (5.73)$$

Dimensional considerations suggest that $A(\mathbf{q}) = \omega_p^2(\mathbf{q})$. In addition, for $q = 0$ it turns out that $\omega_p^2(0) = 16\pi e^2 \Delta E |\langle \psi_1 | \hat{\mathbf{q}} \cdot \mathbf{r} | \psi_2 \rangle|^2 / \hbar^2$, whereas for $q \rightarrow \infty$ one has $\omega_p^2(\infty) = 0$, if the functions ψ_1 and ψ_2 are integrable. Finally, setting $\tau^{-1} = 2\eta/\omega$ the function $\epsilon_L(q, \omega)$ acquires the same form as in (5.12).

Q18. Consider the function $\epsilon_T(\mathbf{q}, \omega)$ in the $\mathbf{q} \rightarrow 0$ limit. Is the result of Q17 valid?

Answer - Yes, it is: indeed $\epsilon_T(\mathbf{q}, \omega)$ coincides with $\epsilon_L(\mathbf{q}, \omega)$ as long as $\mathbf{q} \rightarrow 0$.

Q19. How does the probability in (5.55) change if spinless bosons are considered instead of fermions? and the dissipated power?

Answer - Besides the obvious absence of a factor of 2 due to the spin, the transition probability would be driven by factors $f(1+f)$, with $f = [1 + e^{E_i/(k_b T)}]^{-1}$

the Bose occupation probability of the given energy state with energy E_i : the particles being bosons, the probability of getting them together in the final state is enhanced instead than depressed, as it does in the fermion case. Repeating the calculation of dissipated power for this case, it is immediately seen that the dissipated power does not change: it is indeed a macroscopic result, that cannot distinguish the statistics of the particles.

5.3.3 Applications of the longitudinal dielectric function

In this section the findings on the longitudinal and transverse dielectric function are applied to metals and semiconductor materials.

5.3.3.1 Metals

In the case of metals, the unperturbed Hamiltonian in (5.52) can be approximated as $H_0 = p^2/(2m)$, that is a free-particle Hamiltonian. The-lowest order quantum calculation of $\epsilon_L(\mathbf{q}, \omega)$ is due to Lindhard [5, 7, 14]. From (5.63) one obtains

$$\epsilon_L(\mathbf{q}, \omega) = 1 + \frac{8\pi e^2}{q^2 V} \sum_{\mathbf{k}} \frac{f(\mathbf{k}) - f(\mathbf{k} + \mathbf{q})}{E(\mathbf{k} + \mathbf{q}) - E(\mathbf{k}) - \hbar\omega - i\eta}, \quad (5.74)$$

with $E(\mathbf{k}) = \hbar^2 k^2/(2m)$ and $f(\mathbf{k})$ the Fermi distribution function for free particles. Appendix 5.9 reports its general properties.

Here, the following limiting behaviors are usefully remarked. In the $\mathbf{q} \rightarrow 0$ limit one has

$$\begin{aligned} \epsilon_L(\mathbf{q} \rightarrow 0, \omega) &= 1 - \frac{16\pi e^2}{q^2 V} \sum_{\mathbf{k}} f(\mathbf{k}) \left[\frac{E(\mathbf{k} + \mathbf{q}) - E(\mathbf{k})}{(\hbar\omega + i\eta)^2} + \frac{[E(\mathbf{k} + \mathbf{q}) - E(\mathbf{k})]^3}{(\hbar\omega + i\eta)^4} \right] = \\ &= 1 - \frac{16\pi e^2}{q^2 V} \sum_{\mathbf{k}} f(\mathbf{k}) \left[\frac{\epsilon_{\mathbf{q}}}{(\hbar\omega + i\eta)^2} + \frac{1}{(\hbar\omega + i\eta)^4} \frac{3\hbar^4 \epsilon_{\mathbf{q}}}{m^2} (\mathbf{k} \cdot \mathbf{q})^2 \right], \end{aligned} \quad (5.75)$$

where $\epsilon_{\mathbf{q}} \equiv \hbar^2 q^2/(2m)$ has been defined for convenience. Since $V^{-1} \sum_{\mathbf{k}} f(\mathbf{k}) = n/2$ and $V^{-1} \sum_{\mathbf{k}} (\hbar k/m)^2 f(\mathbf{k}) = (3/5)n v_f^2/2$ with $v_f = \hbar k_f/m$ the velocity of electrons at the Fermi level, one has

$$\epsilon_L(q \rightarrow 0, \omega) = 1 - \frac{\omega_p^2}{(\omega + i\eta/\hbar)^2} - \frac{3}{5} v_f^2 q^2 \frac{\omega_p^2}{(\omega + i\eta/\hbar)^4}. \quad (5.76)$$

One thus obtains the result (5.38) for the dielectric function of metals, that is

$$\epsilon_L(0, \omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\omega/\tau}, \quad (5.77)$$

with $\tau = 2\eta/\hbar$.

In the limit $q \rightarrow 0$ and $\omega \neq 0$, since $\text{Im}(\epsilon_L(q, \omega)) = 0$ from (5.76) one obtains

$$\epsilon_L(q \rightarrow 0, \omega) = 1 - \frac{\omega_p^2}{\omega^2} - \frac{3}{5} v_f^2 q^2 \frac{\omega_p^2}{\omega^4}. \quad (5.78)$$

This equation provides the plasmon dispersion

$$\omega_{pl} = \omega_p \left(1 + \frac{3}{10} \frac{v_f^2 q^2}{\omega_p^2} \right) \quad (5.79)$$

at higher order than in (5.41).

From (5.65) and (5.77) one finds

$$-\text{Im} \left(\frac{1}{\epsilon_L(0, \omega)} \right) = \frac{\omega}{\tau} \frac{\omega_p^2}{(\omega^2 - \omega_p^2)^2 + \omega^2/\tau^2}, \quad (5.80)$$

indicating that the dissipated power is maximum at $\omega = \omega_p$. This result had been anticipated by reading the experimental data in Fig. 5.5, and is in accord with the phenomenological model.

5.3.3.2 Semiconductors and insulators

In the $q \rightarrow 0$ limit, the longitudinal and transverse dielectric functions coincide. In addition, at $T = 0$ the valence band is fully occupied and the conduction band empty, the top of the valence being separated from the bottom of the conduction band by the energy gap E_g . This suggests the following expression for the imaginary part of the dielectric function:

$$\epsilon_2(\omega) = \frac{\pi}{2} \frac{\omega_p^2}{\omega_{av}} \delta(\omega - \omega_{av}), \quad (5.81)$$

where $\hbar\omega_{av}$ is of the order of E_g , ω_p is the plasma frequency of the electrons and the factor $\pi\omega_p^2/2$ has been introduced to account for the sum rule

$$\int_0^\infty \omega \epsilon_2(\omega) d\omega = \frac{\pi}{2} \omega_p^2 \quad (5.82)$$

demonstrated in Appendix 5.10. From Kramers-Kronig relations the real part is calculated

$$\epsilon_1(\omega) = 1 + \frac{\omega_p^2}{\omega_{av}^2 - \omega^2}. \quad (5.83)$$

Equations (5.81) and (5.83) can be combined into the expression for the longitudinal dielectric function

$$\epsilon_L(\omega) = 1 + \frac{\omega_p^2}{\omega_{av}^2 - (\omega + i\eta)^2}. \quad (5.84)$$

The frequency of the longitudinal mode defined by $\epsilon_L(\omega) = 0$ is:

$$\omega_{pl} = \sqrt{\omega_p^2 + \omega_{av}^2}, \quad (5.85)$$

that turns out to be larger than the plasma frequency ω_p . Similar results have been noticed from other models, as in (5.42). Using the dielectric function (5.84) and the relation (5.65) one obtains

$$-\text{Im}\left(\frac{1}{\epsilon_L(\omega)}\right) = \frac{\pi}{2} \frac{\omega_p^2}{\omega_{pl}} \delta(\omega - \omega_{pl}), \quad (5.86)$$

that is experimentally well verified.

Quick Questions

Q20. A laser is a light amplifier. At least four are the main working concepts. First, stimulated emission of light, occurring when an atom in an excited state is stimulated through sympathetic vibration by a photon to decay to its ground state: a photon is then radiated that is a perfect copy of the stimulating photon. Second, population inversion, that is a suited pumping mechanism moving the atoms from the ground to an excited state such that it cannot immediately decay back to the ground: to this aim, a scheme with at least three levels is needed, with one intermediate state between the ground and the population inversion state, where the atoms preferably move from the highest level before decaying into the ground state. Third, a resonance cavity where the photons can travel many times back and forth and be repeatedly duplicated. Fourth, a mechanism that starts the cavity oscillation, making available the first photons to be duplicated. Portable lasers are those mounted inside pointers, barcode readers, CD and DVD players: they are based on semiconductor diodes emitting amplified, coherent light, named Light-Emitting Diodes or LEDs. Using what you know about semiconductor energy bands, discuss would you apply the four concepts above to make such LEDs to work? [1]

Answer - In a forward biased diode, conduction electrons in the n-type semiconductor flow across the junction and become conduction electrons in the p-type part: the conduction band of the p-type semiconductor can be the excited state,

whereas it has empty valence states. The semiconductor material is to favor radiative transitions. For example, Ga, As, Al, based LEDs are good working, Si-based do not. The junction has to be designed to limit reabsorption of the emitted photon, before it leaves the LED. Notice that the emitted photon energies correspond to the band gap: a typical energy is 1.9 eV, the corresponding laser wavelengths being 650 nm, in the red part of the spectrum. Typical required forward bias are as low as 2 V.

5.3.4 Applications of the transverse dielectric function

5.3.4.1 Semiconductors or insulators: interband electromagnetic transitions

The description of interband electromagnetic transitions requires the transverse dielectric function. At zero temperature, (5.72) implies

$$\epsilon_T(\mathbf{q}, \omega) = 1 + \frac{8\pi e^2}{m^2} \frac{1}{V} \sum'_v \sum''_c \frac{|\langle \psi_c | e^{i\mathbf{q}\cdot\mathbf{r}} \mathbf{e} \cdot \mathbf{p} | \psi_v \rangle|^2}{(E_c - E_v)^2 / \hbar^2} \times \frac{1}{E_c - E_v - \hbar\omega - i\eta}, \quad (5.87)$$

where the sum over the occupied states has been distinguished with a prime from that over the unoccupied ones, labeled with a double prime. In the case of a semiconductor, the involved unoccupied states are the Bloch states $\psi_{c,\mathbf{k}'}$ and $\psi_{v,\mathbf{k}}$ of a conduction and of a valence band with energies E_c and E_v , respectively. Use of the Bloch theorem leads to the conclusion that the matrix element $\langle \psi_{c,\mathbf{k}'} | e^{i\mathbf{q}\cdot\mathbf{r}} \mathbf{e} \cdot \mathbf{p} | \psi_{v,\mathbf{k}} \rangle$ is non-vanishing if

$$\mathbf{k}' = \mathbf{k} + \mathbf{q} + \mathbf{G}, \quad (5.88)$$

with \mathbf{G} a reciprocal lattice vector. The vector \mathbf{q} represents the wavenumber of the electromagnetic radiation: it is easily seen that transitions in the visible frequency range require $q \ll q_B$, with $q_B \simeq 10^8 \text{ cm}^{-1}$ being the size of the Brillouin zone. Expression (5.88) states that electromagnetic transitions between two Bloch states are vertical, that is $\mathbf{k}' \simeq \mathbf{k}$. Therefore, the imaginary part of the dielectric function is evaluated to be

$$\epsilon_{2,T}(\omega) = \frac{8\pi^2 e^2}{m^2 \omega^2} \sum_{v,c} \int \frac{d\mathbf{k}}{(2\pi)^3} |\langle \psi_{c,\mathbf{k}} | \mathbf{e} \cdot \mathbf{p} | \psi_{v,\mathbf{k}} \rangle|^2 \delta(E_c(\mathbf{k}) - E_v(\mathbf{k}) - \hbar\omega). \quad (5.89)$$

Under the simplifying assumption that the matrix element be only slightly dependent on \mathbf{k} , one has

$$\begin{aligned}
\varepsilon_{2,T}(\omega) &= \frac{8\pi^2 e^2}{m^2 \omega^2} \sum_{c,v} |\mathbf{e} \cdot \mathbf{M}_{cv}|^2 \int \frac{d\mathbf{k}}{(2\pi)^3} \delta(E_{c,v}(\mathbf{k}) - \hbar\omega) \\
&= \frac{8\pi^2 e^2}{m^2 \omega^2} \sum_{c,v} |\mathbf{e} \cdot \mathbf{M}_{cv}|^2 \frac{1}{(2\pi)^3} \int_{E_{c,v}(\mathbf{k})=\hbar\omega} \frac{dS}{|\nabla E_{c,v}(\mathbf{k})|}. \quad (5.90)
\end{aligned}$$

Here, $\mathbf{M}_{cv} = \langle \psi_c | \mathbf{p} | \psi_v \rangle$, $E_{c,v}(\mathbf{k}) = E_c(\mathbf{k}) - E_v(\mathbf{k})$, and the last integration domain is the surface defined by $E_{c,v}(\mathbf{k}) = \hbar\omega$. The quantity

$$g_{c,v}(\omega) = \frac{1}{(2\pi)^3} \int_{E_{c,v}(\mathbf{k})=\hbar\omega} \frac{dS}{|\nabla E_{c,v}(\mathbf{k})|} \quad (5.91)$$

represents the joined density of states. In the vicinity of critical points, $\nabla_{\mathbf{k}} E_{c,v}(\mathbf{k}) = 0$ and its behavior with varying $\hbar\omega$ has already been described. Therefore,

Concept

Optical absorption measurements may provide information on the joined density of states.

A direct semiconductor is clearly transparent to radiation with frequency $\omega \leq E_g/\hbar$, but provided that \mathbf{M}_{cv} be non zero for symmetry reasons. In Fig. 5.11 the band structure scheme and the expected absorption coefficient are sketched for a direct semiconductor.

Figs. 5.12 and 5.13 instead display the imaginary part of the dielectric function in Ge and graphite, respectively. In the case of Ge, the experimental data are displayed and seen to be qualitatively reproduced by the theoretical prediction corresponding to a three-dimensional joined density of states. In the case of graphite, a good agreement emerges between the imaginary part of the dielectric function and the two-dimensional density of states. In particular, the logarithmic behavior typical of 2D critical points is evident.

Fig. 5.11 Transverse dielectric function for semiconductors. (a) Schematic band structure for a direct semiconductor. (b) Expected qualitative behavior of the absorption coefficient

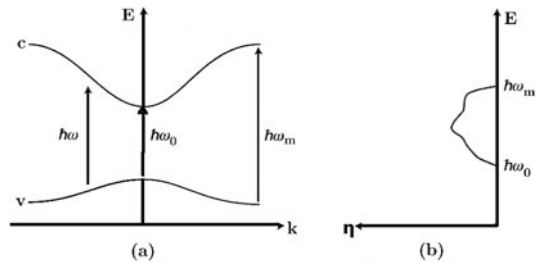


Fig. 5.12 Transverse dielectric function for semiconductors. Imaginary part of the dielectric function in Ge. Solid line: experimental data. Dashed line: theoretical prediction. Reprinted with permission from [11]. Copyright (1962) by the American Physical Society

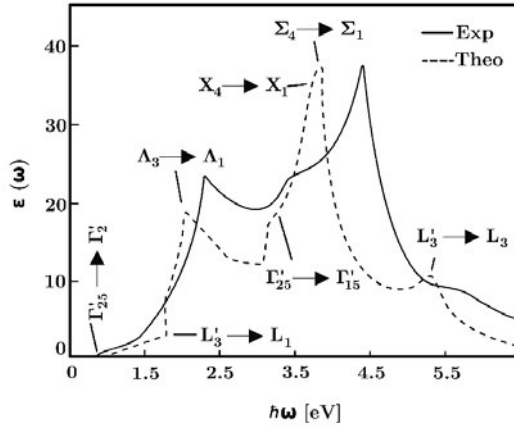
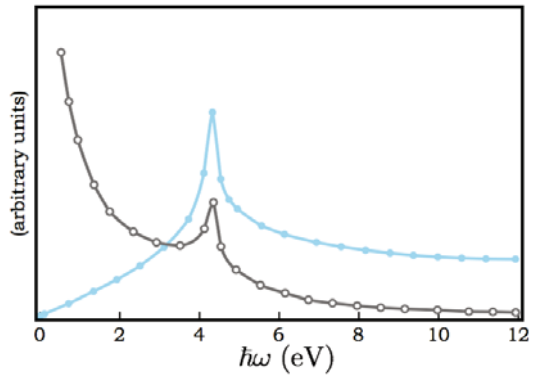


Fig. 5.13 Transverse dielectric function for semiconductors. Imaginary part of the dielectric function in graphite. Open circles: prediction for the dielectric function. Filled circles: joined density of states. The logarithmic singularity typical of 2D critical points is clearly visible. Reprinted with permission from [12, 13]. Copyright (1965) by the American Physical Society



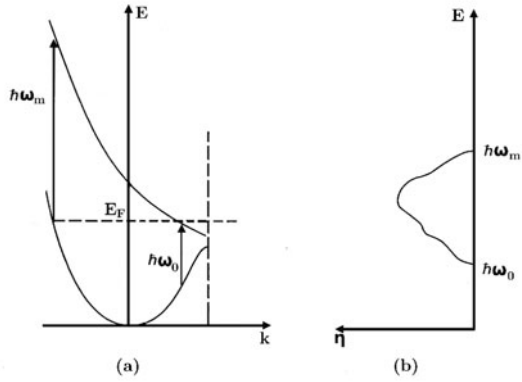
5.3.4.2 Metals: interband electromagnetic transitions

The selection rules (5.88) are again valid for Bloch states in two different bands. They cannot be fulfilled instead for states within the same band, if the frequency of radiation is in the visible range. Fig. 5.14 sketches the interband transitions and the absorption coefficient for a metal with two partially occupied bands.

5.4 Interaction effects

The introduction of scattering mechanisms such as those between electrons and impurities or electrons and phonons has been found in Chap. 4 to be crucial to model the experimental results on electric and thermal conductivity and thermoelectric effect. In this section their relevance to understand the optical properties of materials is discussed. The main message is that correlations, besides exchange, drive a screen-

Fig. 5.14 Transverse dielectric function for metals. (a) Schematic representation of the metal with two partially occupied bands. (b) Sketch of the qualitative behavior predicted for the absorption coefficient



ing of electron-electron, electron-impurity and ion-ion interactions that makes these interactions effectively short ranged. This property is what is needed to understand a number of observations: the optical behavior of materials in response to external electromagnetic fields, the fact that electrons in metals behave as if they were free quasi-particles as within the Drude and Sommerfeld hypothesis, and the way to overcome a number of flaws of the Hartree-Fock model.

The investigation of screening effects is therefore a crucial step here. This task is more easily accomplished by means of the dielectric function formalism. The longitudinal dielectric function has been seen to relate the total electric field inside the material to the acting external electric field. In Appendix 5.7 this statement is shown to be equivalent to state that the inverse longitudinal dielectric function be defined by the ratio

$$\phi(\mathbf{q}, \omega) = \frac{\phi_{\text{ext}}(\mathbf{q}, \omega)}{\epsilon_L(\mathbf{q}, \omega)}, \quad (5.92)$$

of the total $\phi(\mathbf{q}, \omega)$ to the external $\phi_{\text{ext}}(\mathbf{q}, \omega)$ electric potentials. In essence,

Concept

The longitudinal dielectric function measures the capacity of the medium to screen the external potential as a result of the internal correlations and system structure. It determines the frequency of propagation of the longitudinal modes via $\epsilon_L(\mathbf{q}, \omega) = 0$.

In the present section the focus is on the screening effects embodied by the dielectric function. In order to clarify the role of screening, the static $\omega \rightarrow 0$ limit is conveniently addressed in order to isolate screening from the issue of longitudinal mode propagation. One more reason to discuss static screening is that static susceptibilities are related to conserved quantities: this provides sum rules for the exact dielectric function that can be used to validate the approximate models.

Two different strategies are hereafter adopted. In the first, the screening effects resulting from electron-electron and from electron-ions interactions are addressed within two approximations: the classical Thomas-Fermi and the quantum Lindhard models. In the second, the screening operated by electrons on ion-ion interactions is addressed. The specific model approximations here discussed serve to grab the basic concepts and relationships between the different concepts: the general treatment is referred to Chap. 6. The section ends with a discussion on the influence of electron-phonon interactions in characterizing the optical system behavior.

5.4.1 Screening of electron-electron interactions

The key idea of screening is the following. Consider to add a test charge to the system of many electrons. This action originates a different density distribution of electrons around it, due to Coulomb interactions. The induced charge density $\rho_{\text{ind}} = \rho(\mathbf{r}) - \rho_0(\mathbf{r}) = -e(n(\mathbf{r}) - n_0(\mathbf{r}))$, that is the difference between the charge density profile $n(\mathbf{r})$, after the introduction of the test charge $\rho_{\text{ext}}(\mathbf{r})$, and the equilibrium density $n_0(\mathbf{r})$ before that, reflects the system correlations. This is the physical content of (5.92). In the particular static limit under consideration, (5.92) can be cast into the form

$$\phi(\mathbf{q}) = \frac{\phi_{\text{ext}}(\mathbf{q})}{\epsilon_L(0, \mathbf{q})}, \quad (5.93)$$

connecting the total $\phi(\mathbf{q})$ to the external $\phi_{\text{ext}}(\mathbf{q})$ electrostatic potentials. Thus, the calculation of $\epsilon_L(0, \mathbf{q})$ provides a description of screening of the Coulomb interactions. On the other hand, the induced density change can be related to the response function $\chi(\mathbf{q})$ via:

$$e(n(\mathbf{q}) - n_0(\mathbf{q})) \equiv en_{\text{ind}}(\mathbf{q}) = \chi(\mathbf{q})\phi_{\text{ext}}(\mathbf{q}). \quad (5.94)$$

Notice that here $\chi(\mathbf{q})$ represents a density response function, at variance with Sec. 5.2.2.1 where the polarization response was considered. Next, Poisson equations state that

$$\begin{aligned} q^2 \phi_{\text{ext}}(\mathbf{q}) &= -4\pi e n_{\text{ext}}(\mathbf{q}), \\ q^2 \phi(\mathbf{q}) &= -4\pi e n(\mathbf{q}), \end{aligned} \quad (5.95)$$

so that by difference

$$en_{\text{ind}}(\mathbf{q}) = -\frac{q^2}{4\pi e^2} (\phi(\mathbf{q}) - \phi_{\text{ext}}(\mathbf{q})) = \chi(\mathbf{q})\phi_{\text{ext}}(\mathbf{q}), \quad (5.96)$$

or else

$$\phi(\mathbf{q}) = (1 - v_{\mathbf{q}}\chi(\mathbf{q}))\phi_{\text{ext}}(\mathbf{q}), \quad (5.97)$$

in terms of the Fourier transform $v_{\mathbf{q}} = 4\pi/q^2$ of the bare Coulomb potential of the unit charge. Using (5.94), one therefore obtains the expression of the longitudinal dielectric function in terms of the response function, that is:

$$\epsilon_L(\mathbf{q}) = \frac{1}{1 - v_{\mathbf{q}}\chi(\mathbf{q})}, \quad (5.98)$$

or else

$$\frac{1}{\epsilon_L(\mathbf{q})} - 1 = -v_{\mathbf{q}}\chi(\mathbf{q}). \quad (5.99)$$

In the limit of weak interactions, expression (5.98) can be expanded to first order in power series, yielding the lowest order approximation for the longitudinal dielectric function

$$\epsilon_L(\mathbf{q}) = 1 + v_{\mathbf{q}}\chi(\mathbf{q}). \quad (5.100)$$

The calculation of $\chi(\mathbf{q})$ can be easily performed after assuming that the total electrostatic potential $\phi(\mathbf{r})$ is slowly varying. In this case, the quasi-classical approximation can be used and the spatial distributions of electrons, both with and without the external charge are, respectively:

$$\begin{aligned} n(\mathbf{r}) &= \int \frac{d\mathbf{k}}{4\pi^3} \frac{1}{e^{\beta(\epsilon_{\mathbf{k}} - e\phi(\mathbf{r}) - \mu)} + 1} \\ n_0 &= \int \frac{d\mathbf{k}}{4\pi^3} \frac{1}{e^{\beta(\epsilon_{\mathbf{k}} - \mu)} + 1}. \end{aligned} \quad (5.101)$$

Notice that in (5.101) the chemical potential μ is set to be the same in both densities. This could be not necessarily the case. However, it can be demonstrated that the value of μ is indeed the same and that it coincides with that of the unperturbed electronic distribution. Indeed, in the adopted quasi-classical model, the maximum electron energy in the infinitesimal i -th volume is given by $E_m(i) = \hbar^2 k_{f,i}^2 / (2m) - e\phi(\mathbf{r}_i)$, where $k_{f,i} = (3\pi^2 n_i)^{1/3}$ is the usual Fermi wavevector, but calculated at the local electron density n_i . One must have $E_m(i)$ independently of the infinitesimal volume, otherwise the electronic distribution would not any longer be of equilibrium: electrons could indeed migrate from volumes at local higher energy $E_m(i)$ towards volumes at lower local maximum energy, thereby reducing the total energy of the system as required for equilibrium conditions. Therefore, if $E_m(i) = E_m$, in the region where $\phi(\mathbf{r}_i) = 0$ E_m coincides with the Fermi energy, that is the chemical potential at $T = 0$. Since the screening process makes $-e\phi(\mathbf{r})$ a short-range potential, certainly a region exists where it vanishes.

Equations (5.101) thus lead to

$$\begin{aligned} n_{\text{ind}}(\mathbf{r}) &= n_0[\mu + e\phi(\mathbf{r})] - n_0(\mu), \\ \rho_{\text{ind}}(\mathbf{r}) &= -en_{\text{ind}}(\mathbf{r}). \end{aligned} \quad (5.102)$$

Using (5.100), performing a Fourier transform, and expanding to linear order in $e\phi(\mathbf{r})$, leads to the expression for the dielectric function $\epsilon_L(\mathbf{q})$:

$$\varepsilon_L(\mathbf{q}) = 1 + \frac{4\pi e^2}{q^2} \frac{dn_0}{d\mu}. \quad (5.103)$$

In the $T \ll T_F$ limit, $dn_0/d\mu \simeq g(\varepsilon_f)$ and consequently

$$\varepsilon_L(\mathbf{q}) = 1 + \frac{k_0^2}{q^2}, \quad (5.104)$$

with $k_0^2 = 4\pi e^2 g(\varepsilon_f)$. Assuming that the external potential be given by a point charge Q , one has for the total electrostatic potential:

$$\phi(\mathbf{q}) = \frac{\phi_{\text{ext}}(\mathbf{q})}{\varepsilon_L(0, \mathbf{q})} = \frac{4\pi Q}{q^2} \frac{1}{1 + k_0^2/q^2}. \quad (5.105)$$

Therefore,

$$\phi(\mathbf{r}) = \frac{Q}{r} e^{-k_0 r} \quad (5.106)$$

has an exponential decaying behavior on the spatial range $1/k_0$:

Concept

The introduction of interactions between the electrons screens the bare Coulomb potential: the range of the total electrostatic potential $\phi(\mathbf{r})$ originated by the external and the induced charge densities, is effectively shortened on the length scale k_0^{-1} .

Definition

k_0^{-1} is called Thomas-Fermi screening length, and the exponentially decreasing potential (5.106) is of the typical Yukawa form.

The total potential (5.106) describes the interaction between a given electron and the test charge, in the presence of all the other electrons. Within the given approximation, it accounts for the correlations between the charges, except for those related to their statistics. From (5.106) the total charge density $\rho(\mathbf{r})$ can be calculated, along with $\rho_{\text{ind}}(\mathbf{r})$ once the point-charge density $\rho_{\text{ext}}(\mathbf{r})$ is given. To this aim, the Poisson equation $\nabla^2 \phi_{\text{ind}}(\mathbf{r}) = -4\pi \rho_{\text{ind}}(\mathbf{r})$, where $\phi_{\text{ind}}(\mathbf{r}) = \phi(\mathbf{r}) - \phi_{\text{ext}}(\mathbf{r})$, is conveniently used, leading to the remarkable result

$$\int d\mathbf{r} \rho_{\text{ind}}(\mathbf{r}) = -Q. \quad (5.107)$$

That is

Concept

The induced charge acts to completely screen the external charge.

Typical sizes of k_0 can be inferred from (5.104), noticing that

$$\frac{k_0^2}{k_f^2} = \frac{4}{\pi} \frac{1}{k_f a_b} \quad (5.108)$$

with a_b the Bohr radius. Thus, $k_0 \simeq k_f$ in metals and therefore the effective potential $\phi(\mathbf{r})$ has a very short range, of the order of interparticle distances. Assuming the electron-electron interaction as given by (5.106), the free-particle spectrum embodies a self-energy term as found in Chap. 2. However, no singularities in the density of states at the Fermi level appears any longer, healing one of the main flaws of Hartree-Fock theory in comparison with experiments.

Turning to the Lindhard model for the dielectric function, this is derived in Appendix 5.9 and expressed in (5.174). In particular, $\epsilon_L(q, 0)$ can be easily calculated since the imaginary part is zero. One obtains

$$\epsilon_L(q, 0) = 1 + \frac{k_0^2}{q^2} F(x), \quad (5.109)$$

with $k_0^2 = 4\pi e^2 g(\epsilon_f)$ and

$$F(x) = \frac{1}{2} + \frac{1-x^2}{4x} \ln \left| \frac{1+x}{1-x} \right| \quad (5.110)$$

conveniently given in terms of the adimensional variable $x = q/(2k_f)$. In the $q \rightarrow 0$ or $q \ll 2k_f$ limit, $F(0) = 1$ and $\epsilon_L(q, 0)$ gets the Thomas-Fermi form. In the $q \rightarrow \infty$ or $q \gg 2k_f$ limit, one has

$$\epsilon_L(q, 0) = 1 + \frac{4}{3} \frac{k_0^2 k_f^2}{q^4}. \quad (5.111)$$

Thus, the Lindhard $\epsilon_L(q, 0)$ is more rapidly decreasing than the Thomas-Fermi one, though in the region $q \approx 2k_f$ the function $F(x)$ is continuous with a negative and divergent first derivative at $q = 2k_f$.

The charge induced around the introduced test charge Ze can also be calculated, that is

$$\rho_{\text{ind}}(r) = \frac{Ze}{(2\pi)^3} \int \left(\frac{1}{\epsilon_L(q, 0)} - 1 \right) e^{i\mathbf{q} \cdot \mathbf{r}} d\mathbf{q}. \quad (5.112)$$

The property of the function $F(x)$ in Lindhard screening has relevant consequences on the spatial behavior of the induced charge density, as it introduces a step discontinuity in $\epsilon_L(\mathbf{q}, 0)$ at $2k_f$. From the theory of Fourier transforms, it is known that this originates an oscillatory behavior $\cos(2k_f r)$ in real space. The oscillations are then exponentially decaying according to what found after (5.106). The peak positions are signatures of the system density via k_f .

In essence

Concept

As seen while sitting on top of the test charge, the induced charge $\rho_{\text{ind}}(r)$ oscillates in space while rapidly reducing to vanishing values as $r \rightarrow \infty$.

This result justifies the base hypothesis of theory of metals, that is considering the electrons as quasi-free particles.

Quick Questions

Q21. In a semiconductor, electrons are excited from the valence to the conduction band. Electrons are correlated among themselves and so are the holes. However, the range of correlations for the electrons is different from that of the holes. Explain how this might occur.

Answer - The Thomas-Fermi screening length scales as the density of states at the Fermi level, and this scales as the particle mass. Electrons and holes can have different effective masses and therefore their mutual interactions be differently screened.

5.4.2 Screening of ion-ion interactions

Maxwell equations in a material with a dielectric function dispersive in space and time, admit a longitudinal solution with $\tilde{\mathbf{E}}(\mathbf{k}, \omega) \parallel \mathbf{k}$, $\tilde{\mathbf{P}}(\mathbf{k}, \omega) \parallel \mathbf{k}$, $\tilde{\mathbf{D}}(\mathbf{k}, \omega) = 0$, $\tilde{\mathbf{B}}(\mathbf{k}, \omega) = 0$ and dispersion relation $\varepsilon(\mathbf{k}, \omega) = 0$. They also admit a transverse solution with $\tilde{\mathbf{E}}(\mathbf{k}, \omega) \perp \mathbf{k}$, $\tilde{\mathbf{P}}(\mathbf{k}, \omega) \perp \mathbf{k}$, $\omega \tilde{\mathbf{B}}(\mathbf{k}, \omega)/c = \mathbf{k} \times \tilde{\mathbf{E}}(\mathbf{k}, \omega)$ and dispersion relation given by $k^2 c^2 / \omega^2 = \varepsilon(\mathbf{k}, \omega)$. In the Drude-Lorentz model for example, or else in the case of polar semiconductors where scattering mechanisms are neglected, it has been found that the dielectric function has a pole at transverse vibrational frequencies.

In metals and ionic materials the positive electric charges are localized on the ions, that oscillate around their equilibrium positions. Both valence and conduction electrons instead, are characterized by wavefunctions delocalized over the whole crystal. Consider the density fluctuation for an ionic charge and be η the ion displacements independently of the atomic species. An electric field is generated, that is $\mathbf{E} = 4\pi e (\sum_{\mu'} n_{\mu'} Z_{\mu'}) \eta$, with $n_{\mu'}$ the number of ions per unit volume, referring to species μ' and charge $Z_{\mu'} e$. The acceleration acquired by an atom of species μ and charge $Z_{\mu} e$ is $\ddot{\eta}_{\mu}$, and satisfies the equation of motion

$$\ddot{\eta}_{\mu} = -4\pi e^2 \frac{Z_{\mu}}{M_{\mu}} \left(\sum_{\mu'} n_{\mu'} Z_{\mu'} \right) \eta. \quad (5.113)$$

The ionic plasma frequency is thus naturally defined for each atomic species and the average ionic plasma frequency is given by

$$\Omega_p^2 = 4\pi e^2 \frac{1}{r} \left(\sum_{\mu} \frac{Z_{\mu}}{M_{\mu}} \right) \left(\sum_{\mu'} n_{\mu'} Z_{\mu'} \right), \quad (5.114)$$

with r the number of atomic species in the elementary cell. Setting $\sum_{\mu'} n_{\mu'} = n_i$ the number of ions per unit volume, defining Z by means of $Z = (\sum_{\mu'} n_{\mu'} Z_{\mu'}) / n_i$ and M through $Z/M = r^{-1} (\sum_{\mu} Z_{\mu} / M_{\mu})$, the average plasma frequency for ions becomes

$$\Omega_p^2 = \frac{4\pi n_i (Ze)^2}{M}. \quad (5.115)$$

Since

$$\frac{\Omega_p^2}{\omega_p^2} = \frac{Zm}{M}, \quad (5.116)$$

and assuming that the electron density be $n = n_i Z$, it is evident that the ionic plasma frequency is much smaller than the electronic one by a factor $\sqrt{Zm/M}$. This conclusion had already been reached earlier in this Chapter by simpler arguments. On the other hand, the mode under consideration is in all respects represented by an acoustical phonon for the material. As discussed in Chap. 3 and at variance with (5.116), its frequency vanishes in the long-wavelength limit $k \rightarrow 0$. In other words,

Concept

The interactions between ions in the crystal are not long ranged at all, due to the fact that valence electrons are not a rigid system: they can polarize and screen the interactions.

The failure of (5.116) is ultimately due to the fact that ions, having a larger inertia than electrons, are unable to follow their motion. Electrons on the other hand, are influenced by ionic oscillations, though from their viewpoint the latter are almost static. All the above considerations suggest to substitute expression (5.116) with

$$\Omega_i^2(\mathbf{k}) = \frac{\Omega_p^2}{\varepsilon_L(\mathbf{k})}, \quad (5.117)$$

where $\varepsilon_L(\mathbf{k})$ is the electronic dielectric function at $\omega = 0$. In the case of metals, using for example the Thomas-Fermi approximation for $\varepsilon_L(k)$ $\varepsilon_L(k) = 1 + k_0^2/k^2$ one obtains

$$\Omega_i^2(k) = \frac{\Omega_p^2 k^2}{k^2 + k_0^2}.$$

In the $k \ll k_0 \simeq k_F$ limit this becomes

$$\Omega_i(k) = \frac{\Omega_p}{k_0} k = v k \quad (5.118)$$

with

$$v^2 = \frac{\Omega_p^2}{k_0^2} = \frac{Zm}{M} \frac{\omega_p^2}{k_0^2} = \frac{1}{3} Z \frac{m}{M} v_f^2. \quad (5.119)$$

Since m/M is very small and $v_f \simeq 10^8$ cm/s, (5.118) predicts that $v \simeq 10^{-2} v_f \simeq 10^6$ cm/s, that is of the same order of magnitude of sound in solids. In essence, electron charges redistribute themselves around the ions in a crystal in a way to lower the total energy, resulting into an effectively short-ranged interactions between ions.

Notice that this concept has already been implicitly used while developing phonon theory in Chap. 3: there, it was stated that the values of the coefficients $\phi_{m,\mu,i}^{n,v,l}$ are rapidly reducing with $|m-n|$, that is while going far away from site m . In conclusion,

Concept

The medium polarizability as represented by $\epsilon_L(k)$ and dictated by the interactions among electrons or electrons and ions modifies the frequencies of collective excitations for both electrons and ions.

Quick Questions

Q22. It has been found that $\Omega_i(k) \simeq vk$ while $k \rightarrow 0$, within a Thomas-Fermi model for screening. Is this result changed if the more refined Lindhard model is assumed?

Answer - No. The Lindhard and Thomas-Fermi dielectric functions share the same limiting behavior as $k \rightarrow 0$.

5.4.3 Effects of electron-phonon interaction on the optical properties

After electron-electron and ion-ion interactions, this section addresses the role of electron-phonon interactions to shape the optical properties of materials. The interactions with impurities is in fact embodied in the present treatment, since in metals and semiconductors the impurity electronic states can be considered as initial or final states for optical transitions.

Accounting for the presence of phonons poses an additional complexity with respect to the treatment in Secs. 5.4.2-5.4.1: radiation and phonon fields act indeed simultaneously. The strategy to untie this complexity is to first identify a quantum-mechanical expression for the electron-phonon interaction, then to calculate the effects of electron-phonon interactions on electronic transitions to first order in perturbation theory. Eventually, the combined effects of electromagnetic radiation and

of vibrational phonon field are investigated to first and second order in perturbation theory.

5.4.3.1 Electronic transitions induced by phonons

The transition probability per unit time has now to be calculated. To this aim, the unperturbed Hamiltonian is considered to be $H_{el} + H_{ph}$ and the perturbation $\sum_{i,\alpha} H_{e-ph}(\mathbf{r}_i, \mathbf{s}_\alpha)$. The unperturbed states are given by fixing the quantum numbers of the electron Bloch states and the phonon occupation numbers for each branch of normal modes and each wavenumber. This is expressed by

$$|\psi\rangle = |m_1 \mathbf{k}_1 \alpha_1, m_2 \mathbf{k}_2 \alpha_2 \dots\rangle = |n_{s_1} \mathbf{q}_1, n_{s_2} \mathbf{q}_2 \dots\rangle, \quad (5.120)$$

where $|m_1 \mathbf{k}_1 \alpha_1, m_2 \mathbf{k}_2 \alpha_2 \dots\rangle$ is the Slater determinant including Bloch states (m_1, \mathbf{k}_1) with spin α_1 , (m_2, \mathbf{k}_2) with spin $\alpha_2 \dots$. In addition,

$$|n_{s_1} \mathbf{q}_1, n_{s_2} \mathbf{q}_2 \dots\rangle = (b_{s_1}^+(\mathbf{q}_1))^{n_{s_1}} (b_{s_2}^+(\mathbf{q}_2))^{n_{s_2}} \dots |0\rangle. \quad (5.121)$$

Here, $|0\rangle$ is the fundamental state for phonons, on which the operator $b_{s_i}^+(\mathbf{q}_i)$ acts to create a phonon with wavenumber \mathbf{q}_i in branch s_i . To linear order in the displacements, H_{e-ph} can be cast in the form derived in Appendix 3.19:

$$H_{e-ph} = \sum_{i,\mathbf{q},s} \sqrt{\frac{\hbar}{2\omega_s(\mathbf{q})}} (b_{s,-\mathbf{q}}^+ + b_{s,\mathbf{q}}) F(\mathbf{q}, s, \mathbf{r}_i). \quad (5.122)$$

The function $F(\mathbf{q}, s, \mathbf{r}_i)$ is defined in Appendix 3.19, and has to satisfy the relation

$$F(\mathbf{q}, s, \mathbf{r}_i + \mathbf{R}_h) = e^{i\mathbf{q} \cdot \mathbf{R}_h} F(\mathbf{q}, s, \mathbf{r}_i). \quad (5.123)$$

Given $\psi_{m_1, \mathbf{k}_1, \alpha_1, m_2, \mathbf{k}_2, \alpha_2 \dots, n_{s_1}, \mathbf{q}_1, n_{s_2}, \mathbf{q}_2}$ as initial state with energy

$$E_i = \sum_{\mathbf{k}_i, m_i} E_{m_i}(\mathbf{k}_i) + \sum_{s_i, \mathbf{q}_i} \hbar \omega_{s_i, \mathbf{q}_i} \left(n_{s_i, \mathbf{q}_i} + \frac{1}{2} \right) \quad (5.124)$$

and $\psi_{m'_1, \mathbf{k}'_1, \alpha'_1, m'_2, \mathbf{k}'_2, \alpha'_2 \dots, n'_{s'_1}, \mathbf{q}'_1, n'_{s'_2}, \mathbf{q}'_2}$ as the final state with energy

$$E_f = \sum_{\mathbf{k}'_i, m'_i} \varepsilon_{m'_i}(\mathbf{k}'_i) + \sum_{s'_i, \mathbf{q}'_i} \hbar \omega_{s'_i, \mathbf{q}'_i} \left(n'_{s'_i, \mathbf{q}'_i} + \frac{1}{2} \right), \quad (5.125)$$

the transition probability to first perturbative order is

$$P_{if} = \frac{2\pi}{\hbar} |\langle \psi_i | H_{e-ph} | \psi_f \rangle|^2 \delta(E_i - E_f). \quad (5.126)$$

Comparing with the expression (5.122) for H_{e-ph} , one sees that the operator is a sum of products of single-particle and single-phonon operators. Therefore one has

$$\begin{aligned} \langle \psi_i | H_{e-ph} | \psi_f \rangle = \\ = \sum_{\mathbf{q}, s} \langle n_{s1}, \mathbf{q}_1, n_{s2}, \mathbf{q}_2, \dots | \sqrt{\frac{\hbar}{2\omega_s(\mathbf{q})}} (b_{s,-\mathbf{q}}^\dagger + b_{s,\mathbf{q}}) | n_{s'_1}, \mathbf{q}'_1, n_{s'_2}, \mathbf{q}'_2, \dots \rangle \times \\ \langle m_1, \mathbf{k}_1, \alpha_1, m_2, \mathbf{k}_2, \alpha_2, \dots | \sum_i F(\mathbf{q}, s, \mathbf{r}_i) | m'_1, \mathbf{k}'_1, \alpha'_1, m'_2, \mathbf{k}'_2, \alpha'_2, \dots \rangle. \end{aligned} \quad (5.127)$$

Since $\sum_i F(\mathbf{q}, s, \mathbf{r}_i)$ is a sum of single-particle operators, the matrix element between the two Slater determinants is non-vanishing only if they are composed by the same Bloch states but one. This result has been demonstrated in Appendix 2.11 of Chap. 2 for a similar case. When the two Slater determinants coincide, the electronic parts of $E_f - E_i$ do compensate with each other. On the other hand, the matrix element between phonon states is nonzero only if the occupation numbers of the initial state coincide with that of the final one, except for one of them that is to differ by ± 1 . As a result, $E_i - E_f \neq 0$ and therefore the transition is forbidden. To first order then, the Slater determinants referring to initial and final states must differ by one Bloch state. Thus, the matrix element between electronic states is reduced to

$$\langle m_p \mathbf{k}_p \alpha_p | F(\mathbf{q}, s, \mathbf{r}) | m'_p \mathbf{k}'_p \alpha'_p \rangle. \quad (5.128)$$

Since F does not depend on spin, the latter turns out to be nonzero only if $\alpha_p = \alpha'_p$. Appendix 3.19 shows also that the matrix element is nonzero only if

$$\mathbf{q} + \mathbf{k}'_p - \mathbf{k}_p = \mathbf{G}, \quad (5.129)$$

as expected from momentum conservation arguments. The same considerations above hold for the matrix elements between phonon states. Therefore, the transition is allowed only if

$$E_{m_p}(\mathbf{k}_p) - E_{m'_p}(\mathbf{k}_p + \mathbf{q}) = \pm \hbar \omega_s(\mathbf{q}), \quad (5.130)$$

as expected from energy conservation arguments.

In general, typical phonon energies are not sufficient to induce an electronic transition between two different bands. Thus, (5.130) is in practice fulfilled only if

$$E_{m_p}(\mathbf{k}_p) - E_{m_p}(\mathbf{k}_p + \mathbf{q}) = \pm \hbar \omega_s(\mathbf{q}) \quad (5.131)$$

Relations (5.129) and (5.130) provide an univocal determination of \mathbf{q} and $\omega_s(\mathbf{q})$, once \mathbf{k}_p e \mathbf{k}'_p have been given.

Including terms to second order in the displacements, one finds that the above selection rules change into

$$\begin{aligned} \mathbf{q} + \mathbf{q}' + \mathbf{k}'_p - \mathbf{k}_p = \mathbf{G} \\ E_{m_p}(\mathbf{k}_p) - E_{m_p}(\mathbf{k}_p + \mathbf{q} + \mathbf{q}') = \pm \hbar \omega_s(\mathbf{q}) \pm \hbar \omega_{s'}(\mathbf{q}'). \end{aligned} \quad (5.132)$$

In this case a universal determination of \mathbf{q} , \mathbf{q}' , $\omega_s(\mathbf{q})$ and $\omega_{s'}(\mathbf{q}')$ for given \mathbf{k}_p and \mathbf{k}'_p is not possible. In essence,

Concept

Transitions between Bloch states in different bands require the creation and/or destruction of more than one phonon. Under normal conditions the electron-phonon interaction drives electronic wavenumber but not band index changes.

Quick Questions

Q23. A semiconductor has a gap of the order of 1 eV. Estimate how many phonons would be needed to assist an interband transition.

Answer - Considering the average energy of one phonon, something like ten phonons would be needed.

5.4.3.2 Two-photon optical and phonon-assisted transitions

Consider now the combined effects of electromagnetic and vibrational fields. Besides the external electromagnetic perturbation H' , the vibrational field H'' is to be considered as a perturbation. Consider the simplified case where only one phonon mode contributes. The perturbations are represented by

$$\begin{aligned} H' &= \frac{eA_0}{mc} \mathbf{e} \cdot \mathbf{p} e^{i(\eta \cdot \mathbf{r} - \omega t)} + \text{h.c.} \\ H'' &= \sum_{\mathbf{q}} V_p(\mathbf{q}) e^{i(\mathbf{q} \cdot \mathbf{r} - \omega_{\mathbf{q}} t)} + \text{h.c.}, \end{aligned} \quad (5.133)$$

with N the number of cells in the crystal, \mathbf{e} and η the radiation polarization and wavenumber, and $V_p(\mathbf{q})$ the amplitude of electron-phonon interaction: its explicit form $\sum_{\mathbf{R}} V_p(\mathbf{q}, \mathbf{R}) e^{i\mathbf{q} \cdot \mathbf{R}}$ is relevant when actual calculations are to be executed.

Transitions to first perturbative order. Two simultaneous perturbations occur, whose probability amplitude is to first order the sum of the two transition amplitudes, one related to electromagnetic and one to phonon-induced transitions. The initial and final states are products of an electronic times a vibrational state. The square moduli for the two transitions are characterized by the following properties:

Properties

P1. Only the electromagnetic perturbation acts: the first order transition probability is nonzero provided that initial and final phonon states coincide,

and that the initial $|\mathbf{v}\mathbf{k}_1\rangle$ and final $|\mathbf{c}\mathbf{k}_2\rangle$ electronic states satisfy the conditions:

$$\begin{aligned}\mathbf{k}_1 &\simeq \mathbf{k}_2 \\ E_c(\mathbf{k}_1) - E_v(\mathbf{k}_1) &= \pm\hbar\omega,\end{aligned}\tag{5.134}$$

as already calculated in (5.88).

P2. Only the phonon perturbation acts: the transition probability is nonzero provided that

$$\begin{aligned}\mathbf{k}_1 + \mathbf{q} + \mathbf{G} &= \mathbf{k}_2 \\ E_n(\mathbf{k}_1 + \mathbf{q}) - E_n(\mathbf{k}_1) &= \pm\hbar\omega_{\mathbf{q}},\end{aligned}\tag{5.135}$$

as already obtained in (5.129) and (5.131). In this case, the final phonon state has one more or one less phonons than the initial state. Since the phonon energy is typically much smaller than the energy gap, phonon transitions occur when the initial and final electronic states belong to the same band. When $\mathbf{G} = 0$ this corresponds to a normal process, whereas when $\mathbf{G} \neq 0$ to an Umklapp process.

P3. The interference terms between these two types of transitions do not provide any contribution since they are nonzero provided that their energy values are very different.

In conclusion,

Concept

To first order, electromagnetic and phonon-driven perturbations act independently to determine the transition processes.

Transitions to second perturbative order. In this case instead, the two perturbations act in a combined manner. The transition probabilities to the second order in perturbation theory have been calculated in Appendix 3.18 of Chap. 3. The corresponding expression contains the product of two matrix elements of the perturbation, one between the initial and an intermediate state, and a second one between the same intermediate state and the final state. Since the perturbation is the sum of two terms, the transition probability is composed by four terms corresponding to electromagnetic and phonon perturbation only, along with the combination of the two, the latter amounting to two terms. The following properties hold for each of the four cases:

Properties

P1. The electromagnetic perturbation is treated to second order: the evaluation of the matrix elements leads to the following selection rules

$$\begin{aligned} \mathbf{k}' &\simeq \mathbf{k} \\ E_{n'}(\mathbf{k}') &= E_n(\mathbf{k}) \pm \hbar\omega \pm \hbar\omega. \end{aligned} \quad (5.136)$$

According to the first, electronic transitions are again vertical, that is the wavenumbers of the initial and final states coincide. By the second sum rule, the final and initial states differ by absorption or emission of two photons, or by emission (absorption) of one photon and absorption (emission) of one more photon. In a crystal with inversion symmetry, the initial and final states must have the same parity, whereas to first order opposite parity is required by the symmetry properties of (5.133). Finally, in semiconductors with $E_g/2 < \hbar\omega < E_g$, one-photon transitions are forbidden and therefore two-photon transitions are the most likely. Fig. 5.15 displays the absorption coefficient of CdS when two-photon transitions are involved.

P2. The vibrational perturbation is treated to second order: the intermediate transitions each require a ± 1 variation in the number of phonons. The selection rules therefore are:

$$\begin{aligned} \mathbf{k}' &= \mathbf{k} + \mathbf{q} + \mathbf{q}' + \mathbf{G} \\ E_n(\mathbf{k}') &= E_n(\mathbf{k}) \pm \hbar\omega_{\mathbf{q},s} \pm \hbar\omega_{\mathbf{q}',s'}. \end{aligned} \quad (5.137)$$

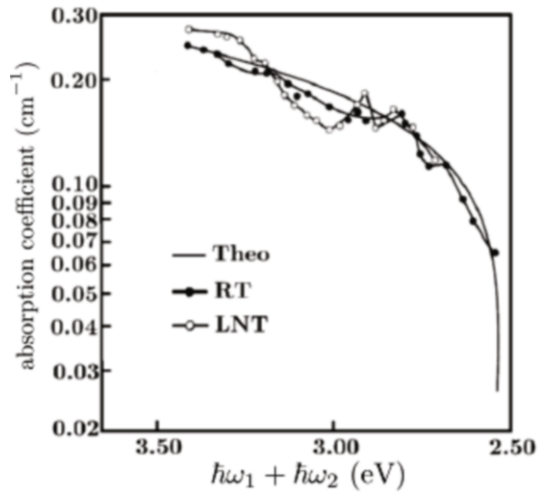
The transition is not vertical, though the energy of the final state is only slightly different from that of the initial state: in essence, an intraband transition occurs. The sum $\mathbf{q} + \mathbf{q}'$ alone is univocally determined and therefore the transitions originate a continuous band. One process with absorption (emission) of two phonons may occur, or else two processes with absorption (emission) of one phonon and emission (absorption) of one more phonon. When $\mathbf{G} = 0$ this corresponds to a normal process, whereas when $\mathbf{G} \neq 0$ one has an Umklapp process.

P3. The electromagnetic and vibrational transitions are treated to second order. The resulting selection rules for the four possible processes are:

$$\begin{aligned} \mathbf{k}' &\simeq \mathbf{k} + \mathbf{q} + \mathbf{G} \\ E_{n'}(\mathbf{k}') &= E_n(\mathbf{k}) \pm \hbar\omega_{\mathbf{q},s} \pm \hbar\omega, \end{aligned} \quad (5.138)$$

with the caveat that the matrix element for the electromagnetic transition is vertical while the vibrational state remains unchanged. That originated by the vibrational perturbation instead, changes the number of phonons by one while the electronic part of the states remains unchanged. The selection rules are not modified if the electromagnetic field acts first and the vibrational one afterward, or viceversa. In this case as well the process with absorption (emission) of both one photon and one phonon may occur, or the process with absorption (emission) of one photon and emission (absorption) of one phonon.

Fig. 5.15 Two-photon transitions. Absorption spectrum for interband direct transitions of CdS with $E_g = 2.44$ eV. Reprinted with permission from [15]. Copyright (1967) by the American Physical Society



In particular, in the case of indirect semiconductors or insulators, it follows that the material absorbs radiations only above a threshold. That is, when the photon energy $\hbar\omega$ is at least

$$\hbar\omega = E_g \pm \hbar\omega_{q,s}, \quad (5.139)$$

with \mathbf{q} the difference between the wavenumber of the valence top and that of the conduction bottom band states. The total number of transitions is found by summing over all electronic states \mathbf{k} and \mathbf{k}' in the vicinity of the valence top and conduction bottom bands, and then over all the phonon wavenumbers \mathbf{q} . Substitution of the quantity $k_b\Theta_D$ to the phonon energies and calculating the number of phonons by means of

$$n_{\mathbf{q}} = \frac{1}{e^{\Theta_D/T} - 1}, \quad (5.140)$$

one finds that the absorption coefficient can be cast into the form

$$\eta(\omega) = A \left[\frac{(\hbar\omega - E_g + k_b\Theta_D)^2}{e^{\Theta_D/T} - 1} + \frac{(\hbar\omega - E_g - k_b\Theta_D)^2}{1 - e^{-\Theta_D/T}} \right], \quad (5.141)$$

with A a slowly-varying function of frequency. The first term represents the indirect transition with annihilation of one phonon, and therefore the energy threshold is $E_g - k_b\Theta_D$. The second term represents the indirect transition with creation of one phonon, and therefore the energy threshold is $E_g + k_b\Theta_D$. In the energy interval $E_g - k_b\Theta_D < \hbar\omega < E_g + k_b\Theta_D$, only the first term exists and thus $\sqrt{\eta(\omega)} \propto \omega$. Fig. 5.16 sketches the behavior of $\sqrt{\eta(\omega)}$ in the case of emission and absorption indirect transitions, as compared to the experimental data for Ge in Fig. 5.17.

Fig. 5.16 Two-photon optical transitions. Sketch of the theoretical predicted behavior of $\sqrt{\eta}$ as a function of frequency corresponding to two temperatures $T_1 < T_2$. Emission and absorption cases are both represented as in the legend. Notice the threshold for the occurrence of the different processes

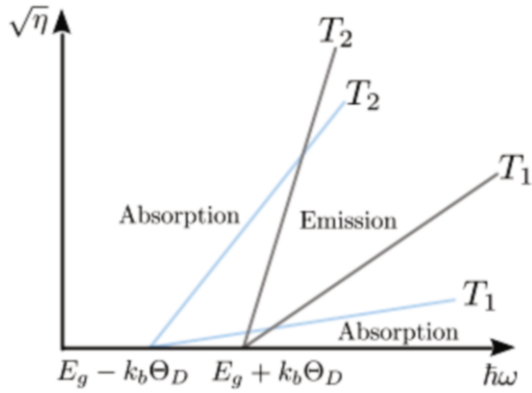
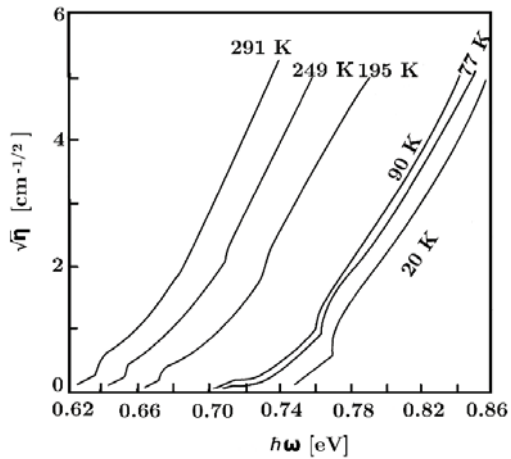


Fig. 5.17 Two-photon optical transitions. Measured $\sqrt{\eta}$ of Ge as a function of frequency, corresponding to different temperatures as in the legend [16]. The qualitative agreement with the theoretical prediction is manifest. Reprinted with permission from [16]. Copyright (1967) by the American Physical Society



5.5 Optical properties of nanostructures

The present section is aimed to illustrate selected relevant results on the optical properties of nanostructures. As examples, the following two sections discuss peculiar and striking optical behaviors in the case of quantum wells and quantum dots, referring to e.g. [17] for the case of quantum wires.

5.5.1 Quantum wells

As seen in Sec. 2.6, quantum wells are realized after layering thin slices of semiconductors, whose widths are on-purpose engineered to drive the wanted quantum-mechanical effects. In Sec. 2.6.1, a simple particle-in-a-box model has been seen to properly describe the main physical properties. In particular, the quantum confinement in the direction perpendicular to the well, say the z direction in Fig. 2.37, determines the energy level quantization according to the characteristic miniband structure with the miniband edges dependent on size. Expression (2.180) shows that the miniband edges are located at the positions $\varepsilon_n^{(z)} = (\hbar^2/2m^*)[n\pi/L]^2$, with L the quantum well width, resulting into the step-like density of states per unit surface:

$$\rho_{2D}(\varepsilon) = \frac{2}{S} \sum_{k_x, k_y, n} \delta \left[\varepsilon - \varepsilon_n^{(z)} - \frac{\hbar^2}{2m^*} (k_x^2 + k_y^2) \right] = \frac{m^*}{\pi \hbar^2} \sum_n \theta \left[\varepsilon - \varepsilon_n^{(z)} \right]. \quad (5.142)$$

Here, $S = L_x L_y$ is the lateral section, $L_x, L_y \gg L$ the lateral sizes of the quantum well. This function is to be compared with the density of states per unit volume of a typical parabolic band in three dimensions, that is $\rho_{3D}(\varepsilon) = (m^*/\pi^2 \hbar^2) \sqrt{(2m^*/\hbar^2)\varepsilon}$ as in Fig. 5.18. As a result, the wavefunctions of an electron confined to move within a

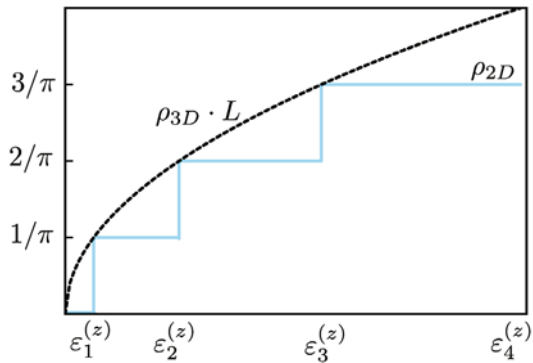


Fig. 5.18 Electronic states in quantum wells. Quantum confinement induces a step-like density of states. For comparison, the density of states of a three-dimensional parabolic band is also shown with a dashed line. Notice that $\rho_{3D}(\varepsilon_n^{(z)}) \cdot L = \rho_{2D}(\varepsilon_n^{(z)})$.

quantum well are labeled by the quantum number n along with the components k_x and k_y of the wave vector in the $x-y$ plane.

As seen in this Chapter, the simplest model to describe absorption processes in bulk semiconductors and in the absence of excitonic effects, is based on the idea that one electron in the valence band absorbs one photon and is thus excited to the conduction band level, its momentum remaining essentially unchanged: the so-called vertical transition as described in Sec. 5.3.4. In the most simple situation, these electron transitions can be considered to occur all with the same strength, though at different energies, as expressed in Eq. (5.90) with the matrix element $\langle \psi_{c,\mathbf{k}} | \mathbf{e} \cdot \mathbf{p} | \psi_{v,\mathbf{k}} \rangle$ approximated to be independent of \mathbf{k} . Therefore, the absorption coefficient onset typically follows the density-of-states dependence on the energy. A measure of the absorption coefficient in such conditions would indeed infer information on the density of states.

In a quantum well, the conservation rule for the momentum along the direction perpendicular to the layers, is to be replaced by a selection rule: to lowest order, only transitions between states sharing the same quantum number in the valence and conduction bands are allowed. Indeed, for sinusoidal standing waves, the overlap integral of the conduction and valence wavefunctions is finite only when the standing waves are similar: in the case of a finite quantum well, that is with a finite boundary potential, this is a good approximation to start with. Because the quantum number n actually labels the allowed values of k_z , this selection rule might also be viewed as a momentum conservation law. This is illustrated in the right panel of Fig. 5.19. On the other hand, as illustrated in the left panel of Fig. 5.19, optical transitions preserve the momentum along the x and y directions: as in the bulk semiconductors, the optical absorption in the xy plane follows the density of states with its characteristic step-like behavior of Fig. 5.18. Because the steps corners touch the density of states in a bulk semiconductor, the ρ_{2D} steps are expected to become increasingly narrower while the well thickness gets larger. In the limit of infinitely large well thickness, the ρ_{2D} steps should merge into the continuous absorption edge of the bulk material.

Fig. 5.19 Optical properties of quantum wells. Left panel: the miniband-to-miniband transitions are governed by the momentum conservation, as in a bulk crystal. Right panel: quantum confinement induces a selection rule, since only transitions from valence to conduction minibands are allowed, which preserve the quantum number n (see text)

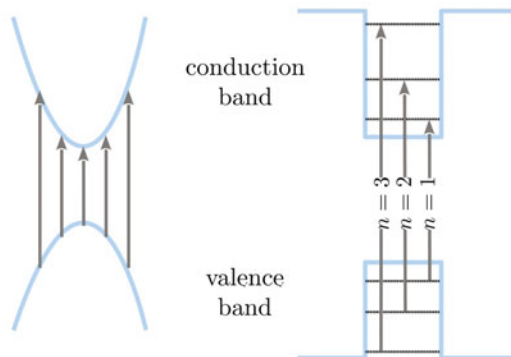


Fig. 5.20 Optical properties of quantum wells. Typical absorption spectrum in a quantum-well structure [18]. The step-like behavior described in the main text and the excitonic peaks are observable

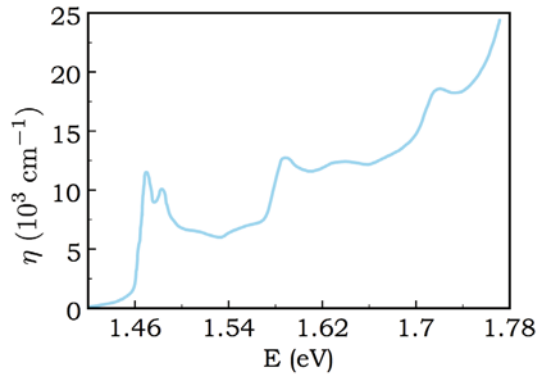


Fig. 5.20 displays a typical absorption spectrum from a quantum well. A series of steps is indeed observable, whose approximate positions are correctly predicted within the simple particle-in-a-box model. Absorption peaks are seen to emerge on top of the step-like structure, that cannot be predicted within this simple model. These are due indeed to excitonic transitions as those described in Sec. 2.6.2. Once excited from the valence into the conduction band, the electron and the corresponding hole, that has left behind, attract each other by Coulomb interaction: this does modify their wavefunctions with respect to the independent, single, particle model and a bound state of the electron-hole pair might form, that is an exciton. The exciton energy is lower than the energy of the free electron and hole, by an amount that corresponds to the binding energy E_b .

Now, at variance with bulk semiconductors, in quantum wells the excitonic effects are visible even at room temperatures, amounting to $\simeq 25$ meV. In fact, incident electromagnetic radiation can thus be absorbed at energies $E_g - E_b$. This process corresponds to an excitonic peak at the exciton ground-state energy, with the 1S wave function peaked at zero relative electron-hole displacement. The amount of absorbed light, and thus the exciton peak strength, are proportional to the probability that electron and hole be found close to each other. The peak strength is also affected by collisions against phonons, producing ionization of the exciton into its originating free electron and hole. In bulk semiconductors, E_b is as small as 1-10 meV and has correspondingly large exciton sizes: in bulk GaAs for example, $E_b \simeq 4$ meV, and the effective Bohr radius is ≈ 300 Å. Therefore, ionization is a likely process and the peak strength is depressed, so that exciton absorption peaks in the bulk can be hardly observed.

Concept

In quantum wells instead, excitons have larger binding energies and smaller sizes than in the bulk, the effect being as neater as the well width decreases.

As a result, ionization after phonon collisions is less likely and the peak strength increases with the probability that electron and hole be close to each other. The signal for excitonic absorption peaks is therefore enhanced, so that eventually quantum well excitons are relatively stronger and better resolved than in bulk semiconductors.

5.5.2 Quantum dots

As detailed in Sec. 2.6.1, quantum confinement in all directions results into a zero-dimensional (0D) nanostructure or quantum dot. Size-dependent optical properties show up independently of the number of surface atoms, that are directly related with the changes in the density of states induced by the quantum confinement of electrons and holes. The atomic-like distribution of electronic levels shown in Fig. 2.38, results in the discrete density of states

$$\rho_{0D}(\varepsilon) = \sum_n \delta(\varepsilon - \varepsilon_n), \quad (5.143)$$

schematically depicted in Fig. 5.21.

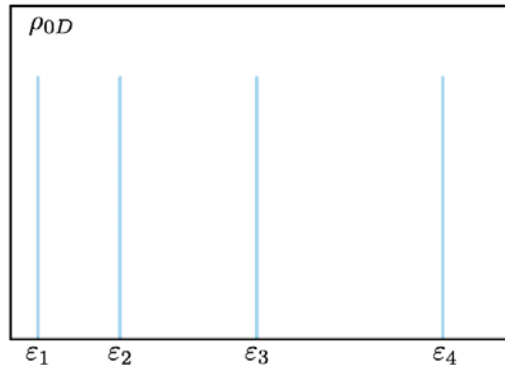


Fig. 5.21 Electronic states in quantum dots. Quantum confinement results in a discrete spectrum, as revealed by the density of states

One of the most striking consequences of zero dimensionality in nanocrystals is that translational symmetry is broken in all three dimensions. At variance indeed with bulk solid, where a continuous distribution of transitions takes place as in the left panel of Fig. 5.19 and in (5.90), only a finite number of electrons and holes can here be created within the same nanocrystal. Therefore, the electron wave vector \mathbf{k} is no longer a good quantum number and for a 0D system the atomic-like states can be considered to arise from the superposition of bulk states with a concurrent dramatic increase of the transition probability.

The typical quantum origin of such behaviour can be cast in terms of the well known uncertainty principle on the definition of position and momentum in quantum systems:

$$\Delta p \cdot \Delta x \geq \hbar/2.$$

For a free particle or a particle moving in a periodic potential, both energy and momentum $\hbar k$ can be precisely defined, while the position cannot. As the particle gets confined, an upper limit to the position uncertainty Δx appears, that is at most equal to the nanostructure size. As a consequence, the momentum cannot be any longer precisely determined. To conclude,

Concept

The transition probability is pinned over sharp transitions, rather than being distributed over a continuum of states as for bulk semiconductors.

This is illustrated in Fig. 5.22, where a schematic view of the electronic transitions between quantum confined states is shown in the left panel. The qualitative behavior of the resulting absorption coefficient is reported in the right panel of the same figure, and compared with the bulk absorption spectrum. The transition from a “discrete” to a “continuum” nature of the spectrum appears evident.

In addition,

Concept

In zero-dimensional nanostructures which the strong-confinement regime applies to, as described in Sec. 2.6.2, the quantum confinement effect can be described using the particle-in-the-box model. The absorption and emission wavelengths depend on the quantum dot radius and geometry. Thus, they can be engineered by controlling the nanostructure shape and size.

This is illustrated in Fig. 5.23, where the absorption spectra of CdSe quantum dots with increasing radii are reported, showing the typical red-shift of the absorption onset on increasing the quantum dot size.

5.6 Experimental methods to determine optical properties

Almost all the experimental methods suited to determine the optical properties of crystals and nanostructures have been basically described in Chaps. 1-3, especially in Secs. 1.6, 2.9, and 3.12. In essence,

Concept

All the spectroscopic methods based on absorption of light, elastic or inelastic scattering of light with appropriate energy and wavelength, and luminescence, are to be considered.

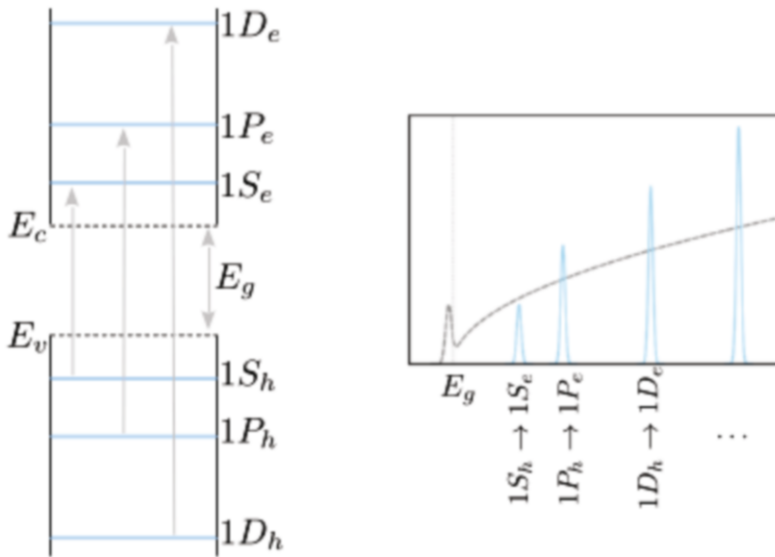


Fig. 5.22 Electronic states in quantum dots. Electronic transitions between single-particle states (left panel) result in a discrete absorption spectrum (right panel). The dashed line in the right panel represents the typical bulk absorption spectrum. The bulk band edges are labeled as E_v and E_c for valence and conduction bands, respectively, and $E_g = E_c - E_v$. The discrete electronic levels arising from quantum confinement are labeled as in Fig. 2.38. The subscripts e and h identify electron (or conduction-like) and hole (or valence-like) states

A few specific remarks are useful though, and discussed in the following.

First, optical transitions from valence to conduction band states in crystals or from the highest occupied to the lowest empty states in nanostructures, require the use of radiation with longer wavelength than the X rays, in particular ultraviolet and visible light with $\lambda \approx 10^{-6} - 10^{-5}$ cm. To first perturbative order, possible processes are absorption or emission of one photon: the occurrence of such transitions requires that the initial and final states have opposite parity whenever the crystal possesses an inversion symmetry. To second perturbative order, scattering processes occur that involve: absorption (emission) of two photons or the absorption (emission) of one photon and the emission (absorption) of one more photon. At variance with the first-order transition, the initial and final states must have the same parity whenever the system has an inversion symmetry. Radiation originated by synchrotron or by solid-state lasers make available tunable radiation sources with high intensity, so that even less-likely transitions can be more easily observed. Though at different energies and wavelengths of the involved light, these processes have been illustrated in Chap. 1-2.

The combined action of radiation and vibrational fields can also originate phonon-assisted transitions, in which the same initial and final state selection rule breaks down. In phonon-assisted transitions to second order, possible processes are either absorption (emission) of one photons and one phonon or the absorption (emission)

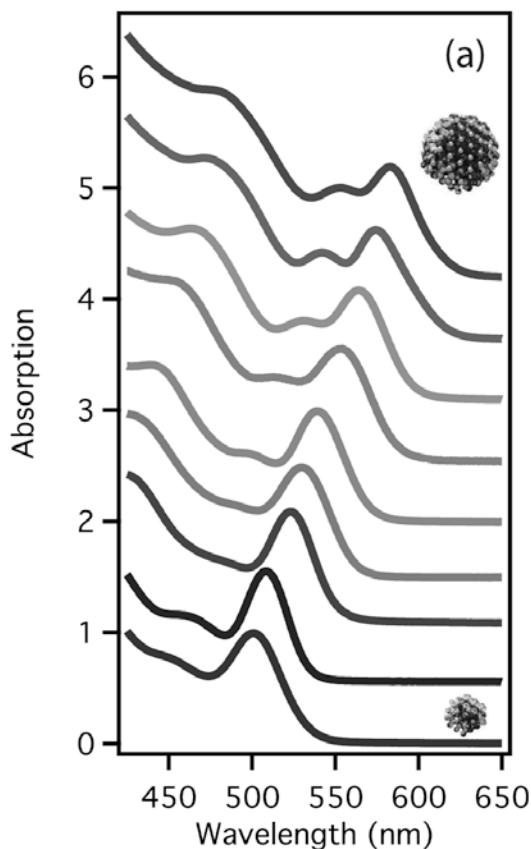


Fig. 5.23 Size-dependent properties in quantum dots. The absorption spectra of colloidal CdSe quantum dots with radii ranging from $R=1.18$ nm (bottom curve) to $R=1.99$ nm (top curve) [19]. The position of the lowest-energy excitonic peak and its shift as a function of the nanocrystal radius is clearly visible

of one photon and the emission (absorption) of one phonon. The involved phonon can be either optical or acoustical. The former case corresponds to Raman and the latter to Brillouin scattering, as illustrated in Chap. 3.

The processes investigated in the present Chapter lead to the idea that

Concept

A measurement of the reflectivity coefficient is one additional method to infer the optical properties.

In metals, electromagnetic radiation is reflected without any absorption for frequencies $\omega\tau \gg 1$ below the plasma frequency: thus below the violet or ultraviolet region of the light spectrum. In insulators, full reflection with no absorption occurs instead for radiation frequencies between ω_T and ω_L : thus in the infrared region of the light spectrum. The case of superconductors has been illustrated in Chap. 2, when discussing the consequences of the existence of a gap in the single-particle excitation spectrum.

Finally, the theoretical framework built up in the present Chapter leads to the idea that

Concept

Measurements of the absorption of light in a wide range of frequencies provide information on the dispersive properties of the medium.

Indeed, it has been shown that the real part of the dielectric function drives the propagation of light in the material, while the imaginary part the dissipative processes. Such quantities are connected by the Kramers-Kronig integral relations: through these relations, the real part of the dielectric function can be calculated from the knowledge of the imaginary part at all frequencies and viceversa. Since real experiments can hardly span the whole frequency range, sum rules may come at hand for cross-checking operations.

Summary: concepts, tools, and procedures to know

Concepts and Tools

- **Weakly excited systems in matter behave as if they were a set of independent damped harmonic oscillators:** the frequency ω_{ij} is related to the energy needed to excite the state and the relaxation time τ describes the state lifetime, hampered by interactions with the surroundings.
- **The average power dissipated by an external force acting on a medium provides information on the induced fluctuations, in particular on frequencies and damping of the modes in the system, or else its excitations.** The link is provided by the imaginary part of the response function, that is in turn connected to the dielectric function. This is how most measurements do work.
- **The dielectric function $\varepsilon(\omega) = E_{ex}(\omega)/E(\omega)$, essentially measures the capacity of the medium of screening the acting external field.**
- The real part of the dielectric function drives the propagation of radiation and the imaginary part determines absorption phenomena.
- Propagation and absorption behaviors are related to each other. Mathematically, through the Kramers-Kronig relations.
- The static dielectric constant $\rightarrow 0$ while the absorption is significant only at $\omega \rightarrow \infty$: static dielectric constants are indeed very large in metals and progressively smaller in semiconductors and insulators.
- Materials able to absorb light at a finite characteristic frequency are completely unavailable to propagate radiation with frequencies lying in the range (ω_0, ω_L) ,

the radiation being therefore totally reflected. The value of ω_L is fixed by the Lyddane-Sachs-Teller relation and larger values of the static dielectric function correspond to smaller ω_0 .

- Systems of many electrons can be characterized by collective long-wavelength modes at the plasma frequency $\omega_p = \sqrt{4\pi n e^2 / m^*}$, corresponding to plasmon excitations.
- In metals, radiations with frequencies $\omega > \omega_p$ easily propagate with negligible absorption, whereas radiation with frequencies $\omega < \omega_p$ do not propagate and are significantly absorbed. If $\tau = 0$ the radiation is completely reflected.
- Concerning absorption and emission, in metals one expects $\hbar\omega \rightarrow (E - \epsilon_f) \rightarrow 0$ since states are immediately available close to the Fermi level, while in semiconductors and insulators one expects that $\hbar\omega > \hbar\omega_0 = E_g$ be larger than the threshold given by the energy gap, progressively increasing while passing from semiconductors to insulators.
- In polar materials, two branches characterize the dispersion relation of light: longitudinal and transverse modes are coupled at intermediate k -range and decouple in the $k \rightarrow 0$ and $k \rightarrow \infty$ limits, where they evolve towards the bare longitudinal and transverse optical modes. This behavior is general, occurring whenever two discrete levels are coupled. In between ω_T and ω_L , a polar material is characterized by full reflectivity.
- Optical absorption measurements may provide information on the joined density of states.
- **The longitudinal dielectric function measures the capacity of the medium to screen the external potential as a result of the internal correlations and system structure, and to determine the propagation of the longitudinal modes by means of $\epsilon_L(\mathbf{q}, \omega) = 0$.**
- Interactions between the electrons screen the bare Coulomb potential, making it short ranged, on the length scale k_0^{-1} .
- In metals, the induced charge acts to completely screen the external charge.
- As seen while sitting on top of the test charge, in metals the induced charge $\rho_{\text{ind}}(r)$ oscillates in space while rapidly reducing to vanishing values as $r \rightarrow \infty$.
- The interactions between ions in the crystal are short ranged, due to the fact that valence electrons are not a rigid system: they can polarize and screen the interactions.
- The medium polarizability as dictated by the interactions among electrons or electrons and ions modifies the frequencies of collective plasma excitations for both electrons and ions.
- **Electromagnetic transitions between Bloch states require that the wavenumbers of the initial and final states are the same.**
- Transitions between Bloch states in different bands require the creation and/or destruction of more than one phonon: electron-phonon interaction typically drives electronic wavenumber but not band index changes.

- To first order, electromagnetic and phonon-driven perturbations act independently to determine the transition processes. To second order, the two are combined leading to the occurrence of absorption or emission transitions where either two phonons, or two photons, or one phonon and one photon are involved.
- In quantum wells, excitons have larger binding energies and smaller sizes than in the bulk, thus ionization after phonon collisions is depressed and excitonic peaks are enhanced in strength and can be resolved, with the probability that electron and hole be close to each other.
- The transition probability in 0D is pinned over sharp transitions, rather than being distributed over a continuum of states as for bulk semiconductors. The absorption and emission wavelengths can be engineered by controlling the nanostructure shape and size.
- **All the spectroscopic methods based on absorption of light, elastic or inelastic scattering of light with appropriate energy and wavelength, and luminescence, are to be considered.** Optical transitions from valence to conduction band states require radiation with longer wavelength than the X rays, in particular ultraviolet and visible light.
- The combined action of radiation and vibrational fields can also originate phonon-assisted transitions, in which the same initial and final state selection rule breaks down. In phonon-assisted transitions to second order, possible processes are either absorption (emission) of one photons and one phonon or the absorption (emission) of one photon and the emission (absorption) of one phonon. The former case corresponds to Raman and the latter to Brillouin scattering.
- Measurements of the reflectivity or of light absorption are additional methods to infer the optical properties. To this aim, it can be useful one or more of the following connected quantities: Fresnel relations, complex refractive index and its relation with the dielectric function, the absorptive nature of the imaginary part of the dielectric function, the Kramers-Kronig relations relating the real part of the dielectric function and, if applicable, the Lyddane-Sachs-Teller relation. Since real experiments can hardly span the whole frequency range, sum rules may come at hand for cross-checking operations.

Procedures

- Determination of the complex and frequency-dependent dielectric function and related polarizability and conductivity response, along with the material behavior with respect to absorption and propagation, as well as refraction, reflection, transmission.
-

5.7 Appendix. Definition of $\varepsilon(\mathbf{k}, \omega)$

For small perturbations and thus safely to linear order, the general equations yielding the polarization $\mathbf{P}(\mathbf{r}, t)$ and the current density $\mathbf{J}(\mathbf{r}, t)$ in response to the application of a total time-dependent electric field are:

$$\begin{aligned}\mathbf{P}(\mathbf{r}, t) &= \int d\mathbf{r}' \int_{-\infty}^t dt' \chi_p(\mathbf{r}, \mathbf{r}', t, t') \mathbf{E}(\mathbf{r}', t') \\ \mathbf{J}(\mathbf{r}, t) &= \int d\mathbf{r}' \int_{-\infty}^t dt' \sigma(\mathbf{r}, \mathbf{r}', t, t') \mathbf{E}(\mathbf{r}', t').\end{aligned}\quad (5.144)$$

Here, $\chi_p(\mathbf{r}, \mathbf{r}', t, t')$ and $\sigma(\mathbf{r}, \mathbf{r}', t, t')$ are the corresponding response functions characterizing the material. Expression (5.144) assumes that the magnetic part of the acting electromagnetic field bears negligible contributions. In essence, the polarization and the current at position \mathbf{r} and time t depend on the action of the total field at times earlier than t and at the remaining space positions with respect to those where the physical quantity is observed and the measurement performed. This is a general manifestation of causality originated by the principle that the speed of the propagating signal cannot exceed the speed of light. The time dependence of these response functions is of the form $t - t'$, since the effects at time $t + \Delta t$ due to the field acting at time $t' + \Delta t$ are to be the same as those corresponding to times t and t' , respectively. If the system can be considered as spatially homogeneous on the scale of spatial variations of $\mathbf{E}(\mathbf{r}, t)$, the dependence on \mathbf{r} and \mathbf{r}' is as well of the form $\mathbf{r} - \mathbf{r}'$, because of translational invariance. In this special case, equations (5.144) can be cast in the form

$$\begin{aligned}\mathbf{P}(\mathbf{r}, t) &= \int d\mathbf{r}' \int_{-\infty}^t dt' \chi_p(\mathbf{r} - \mathbf{r}', t - t') \mathbf{E}(\mathbf{r}', t') \\ &= \int d\mathbf{r}' \int_{-\infty}^{\infty} dt' \chi_p(\mathbf{r} - \mathbf{r}', t - t') \mathbf{E}(\mathbf{r}', t') \theta(t - t') \\ \mathbf{J}(\mathbf{r}, t) &= \int d\mathbf{r}' \int_{-\infty}^{\infty} dt' \sigma(\mathbf{r} - \mathbf{r}', t - t') \mathbf{E}(\mathbf{r}', t') \\ &= \int d\mathbf{r}' \int_{-\infty}^{\infty} dt' \sigma(\mathbf{r} - \mathbf{r}', t - t') \theta(t - t') \mathbf{E}(\mathbf{r}', t').\end{aligned}\quad (5.145)$$

Here,

$$\theta(t) = \begin{cases} 1 & \text{for } t > 0 \\ 0 & \text{for } t < 0 \end{cases}$$

is the Heaviside step function. The above equations represent convolution integrals in both space and time variables, so that their Fourier transform is particularly simple-looking

$$\begin{aligned}\mathbf{P}(\mathbf{k}, \omega) &= \chi_p(\mathbf{k}, \omega) \mathbf{E}(\mathbf{k}, \omega) \\ \mathbf{J}(\mathbf{k}, \omega) &= \sigma(\mathbf{k}, \omega) \mathbf{E}(\mathbf{k}, \omega),\end{aligned}\quad (5.146)$$

in terms of the Fourier-transformed functions $\mathbf{P}(\mathbf{k}, \omega)$, $\mathbf{J}(\mathbf{k}, \omega)$, $\mathbf{E}(\mathbf{k}, \omega)$, $\chi_p(\mathbf{k}, \omega)$ and $\sigma(\mathbf{k}, \omega)$ $\mathbf{P}(\mathbf{r}, t)$, $\mathbf{J}(\mathbf{r}, t)$, $\mathbf{E}(\mathbf{r}, t)$, $\chi(\mathbf{r}, t)\theta(t)$ and $\sigma(\mathbf{r}, t)\theta(t)$, respectively. In order to simplify the notation, the choice has been made to use the same symbols for both Fourier-conjugated quantities, referring to their arguments (\mathbf{r}, t) or (\mathbf{k}, ω) to infer whether the variable is intended to be in space-time or in wavevector-frequency domains.

In conclusion, the equations for the polarization and the current get an algebraic form in the wavenumber and frequency domain. The same response functions in confined crystal structures still conserve the $t - t'$ dependence, but not the $\mathbf{r} - \mathbf{r}'$ one. In this case, the equations become

$$\begin{aligned}\mathbf{P}(\mathbf{r}, t) &= \int d\mathbf{r}' \int_{-\infty}^{\infty} dt \chi_p(\mathbf{r}, \mathbf{r}', t - t') \theta(t - t') \mathbf{E}(\mathbf{r}', t') \\ \mathbf{J}(\mathbf{r}, t) &= \int d\mathbf{r}' \int_{-\infty}^{\infty} dt \sigma(\mathbf{r}, \mathbf{r}', t - t') \theta(t - t') \mathbf{E}(\mathbf{r}', t').\end{aligned}\quad (5.147)$$

Equations (5.146) express the following useful properties of $\chi_p(\mathbf{k}, \omega)$ and $\sigma(\mathbf{k}, \omega)$:

Properties

P1. If $\mathbf{E}(\mathbf{r}', t')$ is slowly-varying in both space and time in the whole region where $\chi_p(\mathbf{r} - \mathbf{r}', t - t')$ is non-vanishing, one has:

$$\mathbf{P}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}, t) \int d\mathbf{r}' \int_{-\infty}^{\infty} dt' \chi_p(\mathbf{r} - \mathbf{r}', t - t') \theta(t - t').$$

Therefore,

$$\mathbf{P}(\mathbf{k}, \omega) = \chi_p(\mathbf{k} = 0, \omega = 0) \mathbf{E}(\mathbf{k}, \omega) \quad (5.148)$$

This corresponds to a local approximation in both space and time.

P2. If $\mathbf{E}(\mathbf{r}', t')$ is slowly varying in space only, one has

$$\mathbf{P}(\mathbf{r}, t) = \int_{-\infty}^{\infty} dt' \left[\int d\mathbf{r}' \chi_p(\mathbf{r} - \mathbf{r}', t - t') \right] \mathbf{E}(\mathbf{r}, t') \theta(t - t')$$

and in Fourier transform

$$\mathbf{P}(\mathbf{k}, \omega) = \chi_p(0, \omega) \mathbf{E}(\mathbf{k}, \omega). \quad (5.149)$$

This corresponds to a local approximation in space .

P3. If $\mathbf{E}(\mathbf{r}', t')$ is slowly varying in time only, thus

$$\mathbf{P}(\mathbf{r}, t) = \int d\mathbf{r}' \left[\int_{-\infty}^{\infty} dt' \chi_p(\mathbf{r} - \mathbf{r}', t - t') \theta(t - t') \right] \mathbf{E}(\mathbf{r}', t)$$

and in Fourier transform

$$\mathbf{P}(\mathbf{k}, \omega) = \chi_p(\mathbf{k}, 0) \mathbf{E}(\mathbf{k}, \omega). \quad (5.150)$$

This corresponds to a local approximation in time .

Now, the Maxwell equations in Fourier transform read

$$\begin{aligned} i\mathbf{k} \cdot \mathbf{E}(\mathbf{k}, \omega) &= 4\pi (-i\mathbf{k} \cdot \mathbf{P}(\mathbf{k}, \omega) + \rho_c(\mathbf{k}, \omega) + \rho_{\text{ext}}(\mathbf{k}, \omega)) \\ i\mathbf{k} \cdot \mathbf{B}(\mathbf{k}, \omega) &= 0 \\ i\mathbf{k} \wedge \mathbf{E}(\mathbf{k}, \omega) &= i \frac{\omega}{c} \mathbf{B}(\mathbf{k}, \omega) \\ i\mathbf{k} \wedge \mathbf{B}(\mathbf{k}, \omega) &= -4\pi i\mathbf{k} \wedge \mathbf{M}(\mathbf{k}, \omega) + \frac{4\pi}{c} \mathbf{j}_c(\mathbf{k}, \omega) - i \frac{\omega}{c} \mathbf{E}(\mathbf{k}, \omega) \\ &\quad - 4\pi i \frac{\omega}{c} \mathbf{P}(\mathbf{k}, \omega) + \frac{4\pi}{c} \mathbf{j}_{\text{ext}}(\mathbf{k}, \omega). \end{aligned} \quad (5.151)$$

Here $\mathbf{M}(\mathbf{k}, \omega)$, $\rho_c(\mathbf{k}, \omega)$ and $\mathbf{j}_c(\mathbf{k}, \omega)$ represent the Fourier transform of the magnetization vector, conduction charge density and current density vector. The phenomena are occurring inside a material, therefore the electric displacement and magnetic field vectors are conveniently introduced in the usual manner:

$$\begin{aligned} \mathbf{D}(\mathbf{k}, \omega) &= \mathbf{E}(\mathbf{k}, \omega) + 4\pi \mathbf{P}(\mathbf{k}, \omega) = (1 + 4\pi \chi_p(\mathbf{k}, \omega)) \mathbf{E}(\mathbf{k}, \omega) \\ \mathbf{H}(\mathbf{k}, \omega) &= \mathbf{B}(\mathbf{k}, \omega) + 4\pi \mathbf{M}(\mathbf{k}, \omega) = (1 + 4\pi \chi_m(\mathbf{k}, \omega)) \mathbf{B}(\mathbf{k}, \omega), \end{aligned} \quad (5.152)$$

where the second equality in each line uses the definition of polarization χ_p and magnetic χ_m susceptibilities, relating the polarization \mathbf{P} and magnetization \mathbf{M} response to the total electric and magnetic field. Equations (5.152) define the dielectric functions

$$\begin{aligned} \varepsilon_p(\mathbf{k}, \omega) &= 1 + 4\pi \chi_p(\mathbf{k}, \omega) \\ \varepsilon_m(\mathbf{k}, \omega) &= 1 + 4\pi \chi_m(\mathbf{k}, \omega). \end{aligned} \quad (5.153)$$

The Maxwell equations can now be cast in the form

$$\begin{aligned} i\mathbf{k} \cdot \mathbf{D}(\mathbf{k}, \omega) &= 4\pi (\rho_c(\mathbf{k}, \omega) + \rho_{\text{ext}}(\mathbf{k}, \omega)) \\ i\mathbf{k} \cdot \mathbf{B}(\mathbf{k}, \omega) &= 0 \\ i\mathbf{k} \wedge \mathbf{E}(\mathbf{k}, \omega) &= i \frac{\omega}{c} \mathbf{B}(\mathbf{k}, \omega) \\ i\mathbf{k} \wedge \mathbf{H}(\mathbf{k}, \omega) &= \frac{4\pi}{c} (\mathbf{j}_c(\mathbf{k}, \omega) + \mathbf{j}_{\text{ext}}(\mathbf{k}, \omega)) - i \frac{\omega}{c} \mathbf{D}(\mathbf{k}, \omega). \end{aligned} \quad (5.154)$$

In usual cases, where no exchanges among different types of charges and currents occur, the relation for the conservation of conduction charge can be used:

$$i\mathbf{k} \cdot \mathbf{j}_c(\mathbf{k}, \omega) - i\omega\rho_c(\mathbf{k}, \omega) = 0.$$

Combining the first and last equations (5.154) with (5.146), one obtains

$$\begin{aligned} i(\varepsilon_p(\mathbf{k}, \omega) - \frac{4\pi}{i\omega}\sigma(\mathbf{k}, \omega))\mathbf{k} \cdot \mathbf{E}(\mathbf{k}, \omega) &= 4\pi\rho_{\text{ext}}(\mathbf{k}, \omega) \\ i\mathbf{k} \wedge \mathbf{H}(\mathbf{k}, \omega) &= -i\frac{\omega}{c}(\varepsilon_p(\mathbf{k}, \omega) - \frac{4\pi}{i\omega}\sigma(\mathbf{k}, \omega))\mathbf{E}(\mathbf{k}, \omega) + \frac{4\pi}{c}\mathbf{j}_{\text{ext}}(\mathbf{k}, \omega). \end{aligned} \quad (5.155)$$

Therefore, the total dielectric function $\varepsilon(\mathbf{k}, \omega)$ is naturally defined

$$\varepsilon(\mathbf{k}, \omega) = \varepsilon_p(\mathbf{k}, \omega) - \frac{4\pi}{i\omega}\sigma(\mathbf{k}, \omega). \quad (5.156)$$

In the special case in which the material is non magnetic, the Maxwell equations would read

$$\begin{aligned} \mathbf{k} \cdot [\varepsilon(\mathbf{k}, \omega)\mathbf{E}(\mathbf{k}, \omega) - \mathbf{E}_{\text{ext}}(\mathbf{k}, \omega)] &= 0 \\ \mathbf{k} \cdot \mathbf{B}(\mathbf{k}, \omega) &= 0 \\ \mathbf{k} \wedge \mathbf{E}(\mathbf{k}, \omega) &= \frac{\omega}{c}\mathbf{B}(\mathbf{k}, \omega) \\ \mathbf{k} \wedge \mathbf{B}(\mathbf{k}, \omega) &= -\frac{\omega}{c}\varepsilon(\mathbf{k}, \omega)\mathbf{E}(\mathbf{k}, \omega) + \frac{4\pi}{ic}\mathbf{j}_{\text{ext}}(\mathbf{k}, \omega), \end{aligned} \quad (5.157)$$

whose study requires the knowledge of the complex function $\varepsilon(\mathbf{k}, \omega)$ of real variables \mathbf{k} and ω .

The first equation in (5.157) states that the transverse electric field, namely characterized by the condition $\mathbf{k} \cdot \mathbf{E}(\mathbf{k}, \omega) = \mathbf{k} \cdot \mathbf{E}_{\text{ext}}(\mathbf{k}, \omega) = 0$ is an automatic solution. For longitudinal fields instead, that are characterized by the condition of non-vanishing $\mathbf{k} \cdot \mathbf{E}(\mathbf{k}, \omega)$ and $\mathbf{k} \cdot \mathbf{E}_{\text{ext}}(\mathbf{k}, \omega)$, the total electric field $\mathbf{E}(\mathbf{k}, \omega)$ is related to the external field $\mathbf{E}_{\text{ext}}(\mathbf{k}, \omega)$ by means of

$$\mathbf{E}(\mathbf{k}, \omega) = \frac{\mathbf{E}_{\text{ext}}(\mathbf{k}, \omega)}{\varepsilon(\mathbf{k}, \omega)}. \quad (5.158)$$

In static or slowly-varying conditions, in which the fields can be represented as due to scalar potentials, the relation (5.158) can alternatively be cast into the form

$$\phi(\mathbf{k}, \omega) = \frac{\phi_{\text{ext}}(\mathbf{k}, \omega)}{\varepsilon(\mathbf{k}, \omega)}, \quad (5.159)$$

with $\phi(\mathbf{k}, \omega)$ and $\phi_{\text{ext}}(\mathbf{k}, \omega)$ the potentials corresponding to the total and external electric fields, respectively.

The third and fourth equations in (5.157) are yet to be analyzed. Assuming that $\rho_{\text{ext}}(\mathbf{k}, \omega)$ and $\mathbf{j}_{\text{ext}}(\mathbf{k}, \omega)$ are vanishing, side-by-side vector multiplication times \mathbf{k}

of the third and use of the fourth yields

$$\mathbf{k} \wedge [\mathbf{k} \wedge \mathbf{E}(\mathbf{k}, \omega)] = -[k^2 \mathbf{E}(\mathbf{k}, \omega) - (\mathbf{k} \cdot \mathbf{E}(\mathbf{k}, \omega))\mathbf{k}] = -\frac{\omega^2}{c^2} \varepsilon(\mathbf{k}, \omega) \mathbf{E}(\mathbf{k}, \omega). \quad (5.160)$$

The solutions of (5.160) are characterized by the following properties:

Properties

P1. If $\mathbf{k} \perp \mathbf{E}$, the third and fourth equations in (5.157) imply that $\mathbf{B} \perp \mathbf{k}$ and $\mathbf{B} \perp \mathbf{E}$. From (5.160), the dispersion relation $\omega(\mathbf{k})$ can be derived as implicitly defined by $k^2 = \omega^2 \varepsilon(\mathbf{k}, \omega)/c^2$. One therefore obtains solutions with electromagnetic character, that correspond to propagation in a dispersive medium.

P2. If $\varepsilon(\mathbf{k}, \omega) = 0$, (5.160) implies that $\mathbf{k} \parallel \mathbf{E}$. From the third equation in (5.157), one has that $\mathbf{B}(\mathbf{k}, \omega) = 0$. One or more longitudinal solutions are therefore obtained, with dispersion relation implicitly defined by $\varepsilon(\mathbf{k}, \omega) = 0$.

Therefore, two different classes of solutions have been found, that are transverse and longitudinal in character. The $k \rightarrow 0$ limit properties of the corresponding dielectric functions do coincide, as it has been remarked in the main text: at long wavelengths, longitudinal and transverse modes cannot be distinguished from each other.

5.8 Appendix. Analytic complex functions and differentiation

A function of complex variable is a univocal correspondence relating a set of complex numbers to another set of complex numbers, according to a given rule.

In analogy with real functions of real variables, a function of complex variable in an open ensemble A of the complex plane, can be derived at point z if and only if the limit

$$\lim_{\Delta z \rightarrow 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} = f'(z) \quad (5.161)$$

exists and it is finite. This condition is somehow tighter than in the case of real variables, since the limit $\Delta z \rightarrow 0$ for complex Δz can be performed in infinite different ways. Setting $z = x + iy$, the function $f(z)$ can be represented by a pair of real functions, writing

$$w = f(z) = u(x, y) + iv(x, y).$$

If $f(z)$ can be derivable, one can take $\Delta z = \Delta x$ and $\Delta z = i\Delta y$, respectively. Therefore,

$$f'(z) = \frac{\partial u(x, y)}{\partial x} + i \frac{\partial v(x, y)}{\partial x} = \frac{1}{i} \left(\frac{\partial u(x, y)}{\partial y} + i \frac{\partial v(x, y)}{\partial y} \right),$$

or in more compact form

$$\frac{\partial f}{\partial x} = \frac{1}{i} \frac{\partial f}{\partial y}.$$

Separation of the real and imaginary parts necessarily yields

$$\begin{aligned} \frac{\partial u(x, y)}{\partial x} &= \frac{\partial v(x, y)}{\partial y} \\ \frac{\partial u(x, y)}{\partial y} &= -\frac{\partial v(x, y)}{\partial x}. \end{aligned} \quad (5.162)$$

The condition (5.162) is also sufficient to state that the function is derivable.

A function $f(z)$ that can be derived at each point of the open ensemble A is named an analytical function.

5.8.1 Integrals of complex-variable functions

Consider a complex-variable function $f(z)$, that be non necessarily analytical. Given two points z' and z'' within its definition ensemble connected by a curve C defined by parametric equations $x = x(t)$ and $y = y(t)$, one defines as integral of the complex function $f(z)$ along the curve C the quantity

$$I = \int_{t'}^{t''} [u(x(t), y(t)) + iv(x(t), y(t))] (dx(t) + idy(t)). \quad (5.163)$$

The compact notation

$$I = \oint_{z'}^{z''} f(z) dz$$

is usually adopted.

The function $f(z)$ admits a primitive function $F(z)$ if for each pair of points z' and z'' it occurs that:

$$\oint_{z'}^{z''} f(z) dz = F(z'') - F(z').$$

Setting in this case $F(z) = U(x, y) + iV(x, y)$, (5.163) implies that the forms $u dx - v dy$ and $v dx + u dy$ have primitive functions $U(x, y)$ and $V(x, y)$, and therefore one has that

$$\begin{aligned} \frac{\partial U}{\partial x} &= u, \quad \frac{\partial U}{\partial y} = -v \\ \frac{\partial V}{\partial x} &= v, \quad \frac{\partial V}{\partial y} = u \end{aligned} \quad (5.164)$$

From the first line it follows that

$$\frac{\partial^2 U}{\partial x \partial y} = \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

and from the second line one gets

$$\frac{\partial^2 V}{\partial x \partial y} = \frac{\partial v}{\partial y} = \frac{\partial u}{\partial x}.$$

Therefore, if the function $f(z)$ admits a primitive function $F(z)$, this is also analytical, even if $f(z)$ does not. In particular, if the curve is closed one has

$$\oint_{\gamma'} f(z) dz = 0.$$

One can prove as well that the analytic character of $f(z)$ is a sufficient condition for the existence of a primitive function.

The following theorem is also easily demonstrated, that is named after Cauchy: be $f(z)$ analytic in the open ensemble A , therefore for a closed curve FT enclosed in it one has that

$$\int_{+FT} f(z) dz = 0.$$

The counterclockwise path is conventionally considered to correspond to the positive sign.

Within the same hypothesis as for the Cauchy theorem, consider z_0 to be a point within the domain delimited by the curve $+FT$. Thus, the function $f(z)/(z - z_0)$ is analytical in the domain enclosed by the curve $+FT$ and external to the circle centered at z_0 with radius ρ , as clockwise covered: the Cauchy theorem holds and in the limit $\rho \rightarrow 0$ one has the quite useful expression:

$$f(z_0) = \frac{1}{2\pi i} \int_{+FT} \frac{f(z)}{z - z_0} dz. \quad (5.165)$$

This is the Cauchy formula.

5.8.2 Improper integrals

Consider e.g. the function $|x - x_0|^\alpha$, with $x_0 \in (a, b)$ and $\alpha > 0$. In the following, the integral

$$I = \int_a^b dx \frac{1}{|x - x_0|^\alpha} \quad (5.166)$$

is conveniently calculated. The integrand function is not continuous and is undefined at x_0 . The usual procedure to calculate the integral is not applicable here. One can extend the definition by considering the two integrals

$$\begin{aligned}
I_1(\varepsilon) &= \int_a^{x_0-\varepsilon} dx \frac{1}{(x_0-x)^\alpha} \\
I_2(\varepsilon') &= \int_{x_0+\varepsilon'}^b dx \frac{1}{(x-x_0)^\alpha},
\end{aligned} \tag{5.167}$$

with ε and ε' positive. If $I_1 = \lim_{\varepsilon \rightarrow 0^+} I_1(\varepsilon)$ and $I_2 = \lim_{\varepsilon' \rightarrow 0^+} I_2(\varepsilon')$ do exist and are finite, thus $I = I_1 + I_2$. Since

$$\begin{aligned}
I_1(\varepsilon) &= \frac{1}{1-\alpha} (\varepsilon^{-\alpha+1} - (x_0-a)^{-\alpha+1}) \\
I_2(\varepsilon') &= \frac{1}{1-\alpha} ((b-x_0)^{-\alpha+1} - \varepsilon'^{-\alpha+1}),
\end{aligned} \tag{5.168}$$

the limits of $I_1(\varepsilon)$ and $I_2(\varepsilon')$ exist and are finite if and only if $\alpha < 1$. In conclusion, the integral I can be defined only for $\alpha < 1$.

If a is finite, $\alpha > 0$, and $a > x_0$, the integral

$$K = \int_a^\infty \frac{dx}{(x-x_0)^\alpha} \tag{5.169}$$

is not defined in the usual sense. In this case as well, one can extend the definition by use of the following procedure. Consider the quantity

$$K_1(T) = \int_a^T \frac{dx}{(x-x_0)^\alpha} = \frac{1}{1-\alpha} [(T-x_0)^{-\alpha+1} - (a-x_0)^{-\alpha+1}].$$

If the limit $K_1 = \lim_{T \rightarrow \infty} K_1(T)$ exists and is finite, K_1 is called the value of the integral (5.169). By inspection, this can be the case if $\alpha > 1$.

The case with $\alpha = 1$ is excluded for both integrals (5.166) and (5.169). Examine then the quantity

$$L = \int_a^b dx \frac{1}{x-x_0}. \tag{5.170}$$

One finds

$$\begin{aligned}
L_1(\varepsilon) &= \ln \left| \frac{\varepsilon}{x_0-a} \right| \\
L_2(\varepsilon') &= \ln \left| \frac{b-x_0}{\varepsilon'} \right|,
\end{aligned} \tag{5.171}$$

diverging as long as ε or ε' become vanishingly small. However, setting $\varepsilon' = \varepsilon$ and performing the $\varepsilon \rightarrow 0$ limit of $L_1(\varepsilon) + L_2(\varepsilon)$, one finds that $L = \ln [(b-x_0)/(x_0-a)]$. This value represents the principal value of the integral (5.170) and is labeled as

$$L = P \int_a^b dx \frac{1}{x-x_0}. \tag{5.172}$$

5.9 Appendix. Lindhard dielectric function

From equation (5.74) one finds that:

$$\begin{aligned}\varepsilon_L(\mathbf{q}, \omega) &= 1 + \frac{8\pi e^2}{q^2 V} \sum_{\mathbf{k}} \left[\frac{f(\mathbf{k})}{E(\mathbf{k} + \mathbf{q}) - E(\mathbf{k}) - \hbar\omega - i\eta} \right. \\ &\quad \left. + \frac{f(\mathbf{k})}{E(\mathbf{k} + \mathbf{q}) - E(\mathbf{k}) + \hbar\omega + i\eta} \right] \\ \varepsilon_L(\mathbf{q}, \omega) &= 1 + \frac{16\pi e^2}{q^2 V} \sum_{\mathbf{k}} f(\mathbf{k}) \frac{E(\mathbf{k} + \mathbf{q}) - E(\mathbf{k})}{[E(\mathbf{k} + \mathbf{q}) - E(\mathbf{k})]^2 - (\hbar\omega + i\eta)^2}. \quad (5.173)\end{aligned}$$

A number of algebraic manipulations lead to the expression of $\text{Re}(\varepsilon_L(\mathbf{q}, \omega))$ and $\text{Im}(\varepsilon_L(\mathbf{q}, \omega))$. In particular, one has

$$\begin{aligned}\text{Re}(\varepsilon_L(q, \omega)) &= 1 + \frac{2me^2 k_F}{\pi\hbar^2 q^2} \\ &\quad \frac{me^2}{\pi\hbar^2 q^3} \left[k_F^2 - \left(\frac{q}{2} - \frac{m}{\hbar} \frac{\omega}{q} \right)^2 \right] \ln \left[\frac{2qk_F + q^2 - (2m/\hbar)\omega}{-2qk_F + q^2 - (2m/\hbar)\omega} \right] + \\ &\quad \frac{me^2}{\pi\hbar^2 q^3} \left[k_F^2 - \left(\frac{q}{2} + \frac{m}{\hbar} \frac{\omega}{q} \right)^2 \right] \ln \left[\frac{2qk_F + q^2 + (2m/\hbar)\omega}{-2qk_F + q^2 + (2m/\hbar)\omega} \right], \quad (5.174)\end{aligned}$$

and

$$\text{Im}(\varepsilon_L(q, \omega)) = \begin{cases} \frac{2m^2 e^2}{\hbar^3} \frac{\omega}{q^3} & \text{case a} \\ \frac{me^2}{\hbar^2} \frac{1}{q^3} \left[k_F^2 - \left(\frac{q}{2} - \frac{m}{\hbar} \frac{\omega}{q} \right)^2 \right] & \text{case b} \\ 0 & \text{case c} \end{cases} \quad (5.175)$$

Case (a)-(c) correspond to the following conditions. Case (a): $q < 2k_f$ and $0 \leq (2m/\hbar)\omega \leq 2qk_f - q^2$; case (b): $2qk_f - q^2 \leq (2m/\hbar)\omega \leq 2qk_f + q^2$; case (c): all other q and ω values. In the limit $q \rightarrow 0$ and $\omega \neq 0$, one also sees that $\text{Im}(\varepsilon_L(q, \omega)) = 0$. The same conclusion is reached in the limit $q \neq 0$ and $\omega \rightarrow 0$.

5.10 Appendix. Sum rules and dielectric functions

Consider the Hamiltonian H_0 for a single particle in a self-consistent potential. The following commutation relations hold:

$$[H_0, e^{i\mathbf{q}\cdot\mathbf{r}}] = e^{i\mathbf{q}\cdot\mathbf{r}} \left(\frac{\hbar^2 q^2}{2m} - \frac{\hbar^2}{m} i\mathbf{q} \cdot \nabla \right), \quad (5.176)$$

$$[e^{i\mathbf{q}\cdot\mathbf{r}}, [H_0, e^{i\mathbf{q}\cdot\mathbf{r}}]] = \frac{\hbar^2 q^2}{m}. \quad (5.177)$$

Calculate now the expectation value of the commutator (5.177) onto the state $|\psi_\alpha\rangle$. Inserting the identity $1 = \sum_\beta |\psi_\beta\rangle\langle\psi_\beta|$ in between the two commutators and using the time-reversal invariance of H_0 , one obtains:

$$\sum_\beta |\langle\psi_\beta|e^{i\mathbf{q}\cdot\mathbf{r}}|\psi_\alpha\rangle|^2 (E_\beta - E_\alpha) = \frac{\hbar^2 q^2}{2m}. \quad (5.178)$$

Here, E_β and E_α label the eigenvalues corresponding to the eigenstates $|\psi_\beta\rangle$ and $|\psi_\alpha\rangle$. Multiplication of the above equation by the Fermi distribution function $f(E_\alpha)$, summation over the α index and division by the volume V yields

$$\frac{1}{V} \sum_{\alpha,\beta} f(E_\alpha) |\langle\psi_\beta|e^{i\mathbf{q}\cdot\mathbf{r}}|\psi_\alpha\rangle|^2 (E_\beta - E_\alpha) = \frac{\hbar^2 q^2}{2m} \frac{n}{2}. \quad (5.179)$$

Here, n labels the number of electrons with the two possible spin orientations and $n/2$ the number of occupied orbitals per unit volume. The left-hand-side of (5.179) is related to the imaginary part of the dielectric function (5.63), so that the following result holds

$$\int_0^\infty d\omega \omega \varepsilon_2(q, \omega) = \frac{\pi}{2} \omega_p^2. \quad (5.180)$$

Therefore, the true and exact imaginary dielectric function is ruled so that the weighted sum of it over all frequencies amounts to a fixed value, $\pi\omega_p^2/2$. This is one among the so-called sum rules, that are relations quite helpful in the task of checking the consistency of approximations or of interpreting experimental data.

Problems with solutions

5.1. Calculate the dielectric function of a system with two absorption frequencies ω_1 and ω_2 with, e.g., $\omega_1 < \omega_2$.

Solution. The imaginary part of the dielectric function is

$$\varepsilon_2(\omega) = C_1 \delta(\omega - \omega_1) + C_2 \delta(\omega - \omega_2). \quad (5.181)$$

The sum rule (5.82) and the Kramers-Kronig relations (5.24) imply

$$C_1 \omega_1 + C_2 \omega_2 = \frac{\pi}{2} \omega_p^2, \quad (5.182)$$

$$\varepsilon_1(\omega) = 1 + \frac{2}{\pi} \left(\frac{C_1 \omega_1}{\omega_1^2 - \omega^2} + \frac{C_2 \omega_2}{\omega_2^2 - \omega^2} \right). \quad (5.183)$$

When $\omega = 0$, this yields

$$C_1 \omega_2 + C_2 \omega_1 = \frac{\pi}{2} (\epsilon_1(0) - 1) \omega_1 \omega_2, \quad (5.184)$$

that together with (5.182) leads to the determination of C_1 and C_2 :

$$\begin{aligned} C_1 &= \frac{\pi}{2} \frac{\omega_1}{\omega_2^2 - \omega_1^2} [(\epsilon_1(0) - 1) \omega_2^2 - \omega_p^2] \\ C_2 &= \frac{\pi}{2} \frac{\omega_2}{\omega_2^2 - \omega_1^2} [\omega_p^2 - (\epsilon_1(0) - 1) \omega_1^2]. \end{aligned} \quad (5.185)$$

From the two equations above one has

$$\begin{aligned} (\epsilon_1(0) - 1) \omega_2^2 - \omega_p^2 &> 0 \\ \omega_p^2 - (\epsilon_1(0) - 1) \omega_1^2 &> 0. \end{aligned} \quad (5.186)$$

Substitution of (5.185) into (5.183) yields

$$\epsilon_1(\omega) = 1 + \frac{1}{(\omega_2^2 - \omega^2)(\omega_1^2 - \omega^2)} [(\epsilon_1(0) - 1) \omega_1^2 \omega_2^2 - \omega_p^2 \omega^2]. \quad (5.187)$$

The latter expression combined with (5.186) yields

$$\lim_{\omega \rightarrow \omega_1^-} \epsilon_1(\omega) = +\infty$$

and

$$\lim_{\omega \rightarrow \omega_2^-} \epsilon_1(\omega) = +\infty.$$

Thus, two solutions exist of the equation $\epsilon_1(\omega) = 0$, say ω_- and ω_+ , with $\omega_1^2 < \omega_-^2 < \omega_2^2$ and $\omega_+^2 > \omega_2^2$ given by

$$\omega_{\pm}^2 = \frac{1}{2} \left[\omega_1^2 + \omega_2^2 + \omega_p^2 \pm \sqrt{(\omega_1^2 + \omega_2^2 + \omega_p^2)^2 - 4\epsilon_1(0)\omega_1^2\omega_2^2} \right], \quad (5.188)$$

that are the two longitudinal solutions, ω_- and ω_+ . The Lyddane-Sachs-Teller relation can be cast in the form

$$\frac{\omega_+^2 \omega_-^2}{\omega_1^2 \omega_2^2} = \epsilon_1(0). \quad (5.189)$$

The model has the property that the calculations and the Lyddane-Sachs-Teller relation hold as well when $\omega_-^2 < \omega_1^2$, though $C_1 < 0$. The model with $C_1 < 0$ means that at frequency ω_1 emission occurs instead than absorption.

5.2. Consider the equations of motion (5.36) and (5.4) for electrons and harmonic oscillators. Calculate the dielectric function of the electron-harmonic oscillators system with densities n and N , respectively. Calculate the frequencies of the longitudinal modes.

Solution. Equation (5.37) yields the electric conductivity related to the electron carriers, that is

$$\tilde{\sigma}(\omega) = \frac{e^2 n}{m^* (-i\omega + 1/\tau_{el})}, \quad (5.190)$$

with τ_{el} the characteristic time of relaxation processes. Equation (5.12) yields the dielectric function for the harmonic oscillators:

$$\varepsilon_{os}(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\omega/\tau_{os}}. \quad (5.191)$$

Here, $\omega_p^2 = 4\pi N q^2 / M$ in terms of the charge q and mass M of the oscillators. Application of (5.156) leads to

$$\begin{aligned} \varepsilon(\omega) &= \varepsilon_{os}(\omega) - \frac{4\pi}{i\omega} \tilde{\sigma}(\omega) \\ &= 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\omega/\tau_{os}} - \frac{\omega_p^2}{i\omega(-i\omega + 1/\tau_{el})}, \end{aligned} \quad (5.192)$$

with ω_p the electronic plasma frequency. It has been shown that for $\omega\tau_{el} \gg 1$ and $\omega\tau_{os} \gg 1$, the imaginary parts of the dielectric functions can be neglected and therefore one obtains

$$\varepsilon(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2} - \frac{\omega_p^2}{\omega^2}. \quad (5.193)$$

The equation $\varepsilon(\omega) = 0$ defines the longitudinal solutions, that are:

$$\omega_{\pm}^2 = \frac{1}{2} \left[\omega_0^2 + \omega_p^2 + \omega_p^2 \pm \sqrt{(\omega_0^2 + \omega_p^2 + \omega_p^2)^2 - 4\omega_0^2 \omega_p^2} \right]. \quad (5.194)$$

These can be studied as functions of n . For $n \rightarrow 0$ and $n \rightarrow \infty$ one has, respectively:

$$\begin{aligned} \omega_-^2 &= \frac{\omega_0^2}{\omega_0^2 + \omega_p^2} \omega_p^2 \\ \omega_+^2 &= \omega_0^2 + \omega_p^2 + \frac{\omega_p^2}{\omega_0^2 + \omega_p^2} \omega_p^2, \end{aligned} \quad (5.195)$$

and

$$\begin{aligned} \omega_-^2 &= \omega_0^2 \\ \omega_+^2 &= \omega_p^2 + \omega_p^2. \end{aligned} \quad (5.196)$$

The Lyddane-Sachs-Teller relation is substituted by the following

$$\frac{\omega_+^2 \omega_-^2}{\omega_0^2 \omega_p^2} = 1. \quad (5.197)$$

Considering the dielectric function (5.49), (5.197) is substituted by

$$\frac{\omega_+^2 \omega_-^2}{\omega_T^2 \omega_p^2} = 1. \quad (5.198)$$

5.3. Recall that a complex wavenumber k is related to the dielectric function $\varepsilon(\omega)$ by means of $k^2 = \omega^2 \varepsilon(\omega)/c^2$, that the dielectric function can be decomposed as $\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$, and that the complex index of refraction $\tilde{n}(\omega) = n(\omega) + i\kappa(\omega)$ is defined by means of $\tilde{n}^2(\omega) = \varepsilon(\omega)$. Find the relationship between $n(\omega)$ and $\kappa(\omega)$, and $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$. Introduce then the complex wavenumber $k = k_1 + ik_2$ such that $k = \omega \tilde{n}(\omega)/c$ and assume $\kappa(\omega) \ll n(\omega)$. Determine the relationship between k_1 and k_2 , and $n(\omega)$ and $\kappa(\omega)$ up to first order in $\kappa(\omega)/n(\omega)$. Apply the results to the case of metals in the region where the imaginary part of the polarization-related dielectric function $\varepsilon_p(\omega)$ and the imaginary part of $\sigma(\omega)$ are zero.

Solution. From $\tilde{n}^2(\omega) = \varepsilon(\omega)$ it follows that $n^2(\omega) - \kappa^2(\omega) = \varepsilon_1(\omega)$ and $2n(\omega)\kappa(\omega) = \varepsilon_2(\omega)$. Under the given hypothesis, $\kappa^2(\omega)$ can be neglected in the first relation and therefore $n(\omega) = \sqrt{\varepsilon_1(\omega)}$. From the second relation, it turns out that $\kappa(\omega) = \varepsilon_2(\omega)/(2n(\omega))$. From $k = \omega \tilde{n}(\omega)/c$ it follows that $k_1 = \omega n(\omega)/c$ and $k_2 = \omega \kappa(\omega)/c$, so that the absorption coefficient $\eta = 2\omega \kappa(\omega)/c$ can be cast in the form

$$\eta = \frac{\omega \varepsilon_2}{n(\omega)c}. \quad (5.199)$$

In the case of metals, from the relation

$$\varepsilon(\omega) = \varepsilon_p(\omega) - \frac{4\pi}{i\omega} \sigma(\omega), \quad (5.200)$$

one has that $\varepsilon_1(\omega) = \varepsilon_p(\omega)$ and $\varepsilon_2(\omega) = 4\pi\sigma(\omega)/\omega$. Thus,

$$\begin{aligned} n(\omega) &= \sqrt{\varepsilon_p(\omega)} \\ \eta &= \frac{4\pi\sigma(\omega)}{n(\omega)c}, \end{aligned} \quad (5.201)$$

yielding a direct link between the absorption coefficient and the conductivity in the metal.

5.4. Given

$$\varepsilon_2(\omega) = \begin{cases} C & \text{for } \omega_a < \omega < \omega_b \\ 0 & \text{otherwise} \end{cases}$$

calculate the dielectric function and the longitudinal modes.

Solution. The sum rule for $\varepsilon_2(\omega)$ yields

$$C = \pi \frac{\omega_p^2}{\omega_b^2 - \omega_a^2}$$

and from the Kramers-Kronig relations one obtains

$$\varepsilon_1(\omega) = 1 + \frac{\omega_p^2}{\omega_b^2 - \omega_a^2} \log \left| \frac{\omega_b^2 - \omega^2}{\omega_a^2 - \omega^2} \right|. \quad (5.202)$$

The longitudinal modes are found by solving the equation

$$1 + \frac{\omega_p^2}{\omega_b^2 - \omega_a^2} \log \left| \frac{\omega_b^2 - \omega^2}{\omega_a^2 - \omega^2} \right| = 0,$$

that yields

$$\begin{aligned} \omega_L^2 &= \frac{\omega_b^2 - \omega_a^2 e^{-\alpha}}{1 - e^{-\alpha}} \\ \alpha &= \frac{\omega_b^2 - \omega_a^2}{\omega_p^2}. \end{aligned} \quad (5.203)$$

It can be verified that $\omega_L^2 > \omega_b^2$.

5.5. In a direct semiconductor with energy gap E_g , the expression (5.70) can be used to calculate the imaginary part of the dielectric function in the vicinity of maxima and minima of the energy bands. Explicitly execute this calculation, assuming that the matrix elements be constant and that the transition from a valence to a conduction band state be vertical.

Solution. From (5.70) and assuming that the matrix elements $\langle \psi_j | e^{i\mathbf{q} \cdot \mathbf{r}} e \cdot \mathbf{p} | \psi_i \rangle = M$ be constant, one finds

$$\begin{aligned} \varepsilon_2(\mathbf{q}, \omega) &= \frac{8\pi^2 e^2}{m^2 \omega^2} \frac{1}{V} |M|^2 \frac{V}{(2\pi)^3} \int d\mathbf{k} \delta(E_c(\mathbf{k}) - E_v(\mathbf{k}) - \hbar\omega) \times \\ &\quad (f(E_v(\mathbf{k})) - f(E_c(\mathbf{k}))). \end{aligned} \quad (5.204)$$

At $T = 0$, one has $f(E_v(k)) = 1$ and $f(E_c(k)) = 0$. In addition,

$$\begin{aligned} E_v(\mathbf{k}) &= -\frac{\hbar^2}{2m_v} k^2 \\ E_c(\mathbf{k}) &= E_g + \frac{\hbar^2}{2m_c} k^2, \end{aligned} \quad (5.205)$$

so that the difference

$$E_c(\mathbf{k}) - E_v(\mathbf{k}) = E_g + \frac{\hbar^2}{2\mu} k^2,$$

with $\mu^{-1} = m_c^{-1} + m_v^{-1}$. The calculation is easily performed, since

$$\int d\mathbf{k} \delta(E_c(\mathbf{k}) - E_v(\mathbf{k}) - \hbar\omega) = 2\pi \left(\frac{2\mu}{\hbar^2} \right)^{\frac{3}{2}} \sqrt{\hbar\omega - E_g}.$$

5.6. Determine the solutions of the Maxwell equations that are in the form

$$\begin{aligned} E_x &= A e^{iqx} e^{-Kz} e^{-i\omega t}, \quad E_z = B e^{iqx} e^{-Kz} e^{-i\omega t}, \quad z > 0, \\ E_x &= C e^{iqx} e^{K'z} e^{-i\omega t}, \quad E_z = D e^{iqx} e^{K'z} e^{-i\omega t}, \quad z < 0, \end{aligned}$$

with $E_y = 0$, and K and K' positive constants. A , B , and C are constants to be determined by means of the boundary conditions for the Maxwell equations at the separation interface between a metal with dielectric function $\varepsilon(\omega)$ placed in the region with $z > 0$ and vacuum in the region with $z < 0$.

Solution. The condition $\nabla \cdot \mathbf{E} = 0$ applied in the two regions with $z > 0$ and $z < 0$ implies:

$$\begin{aligned} iqA - KB &= 0 \\ iqC + K'D &= 0. \end{aligned} \quad (5.206)$$

From $\nabla \wedge \mathbf{E} = i\omega c^{-1} \mathbf{B}$, one obtains $B_x = B_z = 0$ for $z > 0$ and $z < 0$. In addition,

$$\begin{aligned} B_y &= \frac{c}{i\omega} (-KA - iqB) e^{iqx} e^{-Kz} e^{-i\omega t} \\ B_y &= \frac{c}{i\omega} (K'C - iqD) e^{iqx} e^{K'z} e^{-i\omega t}. \end{aligned} \quad (5.207)$$

Continuity of B_y , E_x , and D_z at $z = 0$ implies

$$\begin{aligned} K'C + KA &= iq(D - B) \\ A &= C \\ D &= \varepsilon(\omega)B. \end{aligned} \quad (5.208)$$

Substitution of the second and third lines in the first equation yields

$$(K' + K)A = iq(\varepsilon(\omega) - 1)B. \quad (5.209)$$

Substitution of the second relation of (5.206) into (5.209) results into

$$\frac{q^2}{K'} \frac{\varepsilon(\omega) - 1}{\varepsilon(\omega)} = K + K'. \quad (5.210)$$

For simplicity it is assumed that $\nabla \wedge \mathbf{B} = 0$, that is the fields are slowly-varying in time. This implies that

$$\begin{aligned} KA + iqB &= 0 \\ K'C - iqD &= 0. \end{aligned} \quad (5.211)$$

The ratio of the first (second) among the relations (5.206) to the second (first) among the (5.211) leads to $K = K' = q$. After substitution in (5.210) one obtains $\varepsilon(\omega) = -1$, yielding the frequency of the surface plasmon as $\omega_{PS} = \omega_p / \sqrt{2}$.

5.7. Given the dielectric function (5.13), determine the values of ω that make the real part of $\varepsilon(\omega)$ vanish. Is $\varepsilon_2(\omega)$ negligible?

Solution. Setting $\varepsilon_1(\omega) = 0$ one finds the equation

$$x^2 - x(2a_0 + a_p - 1) + a_0(a_0 + a_p) = 0,$$

with $x = (\omega\tau)^2$, $a_0 = (\omega_0\tau)^2$ and $a_p = (\omega_p\tau)^2$. The solutions are

$$x_{\pm} = \frac{1}{2} \left[2a_0 + a_p - 1 \pm \sqrt{(a_p - 1)^2 - 4a_0^2} \right].$$

The first condition for the solutions to be real is $|a_p - 1| > 2a_0$. In the case with $2a_0 + a_p - 1 > 0$, a solution x_- exists only if $2a_0 + a_p - 1 > \sqrt{(a_p - 1)^2 - 4a_0^2}$: this is indeed the case as it can be checked. In the case with $2a_0 + a_p - 1 < 0$, a solution x_- cannot exist and x_+ exists only if $\sqrt{(a_p - 1)^2 - 4a_0^2} > 1 - 2a_0 - a_p$: this is indeed the case as it can be checked.

In conclusion:

- if $a_p > 1$ and $2a_0 + a_p - 1 > 0$, the solution x_- exists as well only if $a_p > 1 + 2a_0$;
- if $a_p < 1$ and $2a_0 + a_p - 1 > 0$ the solution x_- does not exist;
- if $a_p > 1$ and $2a_0 + a_p - 1 < 0$, no solutions may exist;
- if $a_p < 1$ and $2a_0 + a_p - 1 < 0$, only the solution x_+ exists.

Finally, $\varepsilon_2(\omega)$ is not necessarily negligible.

Problems without solutions

5.8. Calculate the energy loss function F that corresponds to the dielectric function

$$\varepsilon(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\omega/\tau}.$$

Study F as function of τ .

5.9. Consider the dielectric function given in Problem 5.8. Assuming that radiation is propagating along the x -axis, calculate $k(\omega)$ and $\omega(k)$ to characterize the space and time imaginary parts.

5.10. Consider the dielectric function (5.38). Draw $\text{Re}(\varepsilon(\omega))$ and $\text{Im}(\varepsilon(\omega))$ as functions of ω and τ .

5.11. Consider the dielectric function of a polar material with the dissipative term taken into account, as in Q13. Calculate the energy loss function.

5.12. Consider the one-dimensional potential $V = [\delta(x - d/2) + \delta(x + d/2)]$. Calculate the probability that the system may execute a transition between the ground and the excited state.

5.13. Given the expression for the electron plasma frequency, find the electron density corresponding to the frequency value of 10^{13} s^{-1} .

References

1. L. Bloomfield: *How Everything Works. Making physics out of ordinary*. John Wiley and Sons, USA (2008)
2. F. Bassani and U.M. Grassano: *Fisica dello Stato Solido*. 1st ed. Bollati Boringhieri, Torino (2000)
3. C. Kittel: *Introduction to Solid State Physics*. 5th ed. John Wiley, Hoboken (2005)
4. N.W. Ashcroft and N. D. Mermin: *Solid-State Physics*. 2nd ed. Holt-Saunders, Tokyo (1976)
5. G. Grosso and G. Pastori Parravicini: *Solid State Physics*. 1st ed. Academic Press, New York (2000)
6. For a nice pedagogical illustration see, e.g. P. C. Martin: *Measurements and Correlation Functions* in *Many-Body Physics*, C. De Witt and R. Balian Eds. Gordon and Breach, New York (1968)
7. G. D. Mahan: *Many-Particle Physics*. 3rd ed. Kluwer Academic, New York (2000)
8. D.L. Greenaway, G. Harbeke, F. Bassani and E. Tosatti, *Phys. Rev.* **178**, 1340 (1969)
9. C.J. Powell and J.B. Swan, *Phys. Rev.* **115**, 869 (1959)
10. H. Ehrenreich, H.R. Philipp, and B. Segal, *Phys. Rev.* **132**, 1918 (1963)
11. D. Burst, H.R. Phillips, and F. Bassani, *Phys. Rev. Lett.* **9**, 94 (1962)
12. F. Bassani and G. Pastori Parravicini, *Il Nuovo Cimento B* **50**, 95 (1967)
13. E.A. Taft and H.R. Philipp, *Phys. Rev.* **138**, A197 (1965)
14. J. Lindhard, *Kgl. Danske Fys. Medd.* **28**, 8 (1954)
15. P.J. Regesburger and E. Panizza, *Phys. Rev. Lett.* **18**, 113 (1967)
16. G.G. MacFarlane, T.P. McLean, J.E. Quarrington, and V. Roberts, *Phys. Rev.* **108**, 1377 (1967)
17. P. Y. Yu and M. Cardona: *Fundamentals of Semiconductors*. Springer, Berlin (1996)
18. D.A.B. Miller, *Opt. Photon. News* **1**, 7 (1990)
19. V.M. Huxter, and G.F. Scholes, *J. Nanophot.* **3**, 032504 (2009)

Chapter 6

Correlations and Density Functional Theory

Abstract The Density Functional Theory (DFT) and its extensions to time-dependent systems and to special types of ground states is formulated, as a powerful framework to perform explicit and accurate calculations for atoms, molecules, crystals and nanosystems and obtain good agreement in comparison with experimental findings. Concepts, methods, and applications are illustrated. The discussion around the DFT represents a framework where many puzzle pieces found in previous Chapters fit together: the approach adopted within approximate models such Thomas-Fermi and Hartree and Hartree-Fock methods, the density and current-density response theories used to understand the transport in Chap. 4 and optical in Chap. 5 behaviors, the crystal, electron and phonon structures investigated in Chaps. 1, 2 and 3.

6.1 Introduction

The inclusion of many-particle correlations to understand the properties of crystals and nanostructures is in principle a formidable task. The Density Function Theory (DFT) and its extensions to time-dependent systems and to special types of ground states represent a powerful key advancement in conceptual, procedural, and factual knowledge. The DFT concept is based on the Hohenberg and Kohn theorem, stating that the ground state energy of a correlated system of many particles in an external potential is a functional of the density profile, and on the Kohn and Sham result, stating that the many-particle system can be mapped onto a suited and fictitious non-interacting system of particles moving in an effective potential. Practical schemes are available to apply these concepts and calculate the predicted behavior of microscopic quantities that can be observed in atomic, molecular, solid-state physics. The DFT predictions are most often found to provide excellent agreement with experimental findings, sometimes in fact even outside the range of validity of the adopted approximation that can be inferred a priori.

The problem of many interacting electrons has so far been treated within the Hartree and Hartree-Fock methods and the Born-Oppenheimer approximation in Chap. 2, the first to be accurate enough to determine the electronic states of atoms, molecules, and crystals [1, 2]. Chap. 5 has introduced the concept of screening to investigate the manner in which correlations affect the optical properties, and the Thomas-Fermi and Lindhard approximations have in particular been introduced to illustrate this issue. The present Chapter restarts from the Hartree-Fock results and the Thomas-Fermi screening approach to calculate the lowest order expressions for kinetic, exchange and correlation energies. The discussion developed around the determination of the ground state energy is used to introduce the Density-Functional Theory method [3, 4]. The connection of DFT to other theories is then discussed. Then, selected applications of DFT are illustrated, with particular attention to the problem of screening in crystals and nanostructures. Finally, the extension of DFT to time-dependent physical processes is briefly illustrated.

6.2 Exchange and correlation in N -particle systems

Consider one among the N interacting particles in the system. The degree of correlation can be measured by the probability of finding a second particle at a given distance. Formalization of this concept may proceed by defining the single-particle and pair densities in terms of the wavefunction $\Psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \cdots \mathbf{r}_N)$ of the system with N particles:

Definition

The single-particle density $n(\mathbf{r})$ is

$$n(\mathbf{r}) = N \int |\Psi(\mathbf{r}, \mathbf{r}_2, \cdots \mathbf{r}_N)|^2 d\mathbf{r}_2 \cdots d\mathbf{r}_N. \quad (6.1)$$

The pair density $n_2(\mathbf{r}, \mathbf{r}')$ is

$$n_2(\mathbf{r}, \mathbf{r}') = \frac{1}{2} N(N-1) \int |\Psi(\mathbf{r}, \mathbf{r}', \mathbf{r}_3 \cdots \mathbf{r}_N)|^2 d\mathbf{r}_3 \cdots d\mathbf{r}_N. \quad (6.2)$$

Alternatively, $\hat{\rho}_1(\mathbf{r}) = \sum_i \delta(\mathbf{r} - \mathbf{r}_i)$, and $\hat{\rho}_2(\mathbf{r}, \mathbf{r}') = 2^{-1} \sum_{i \neq j} \delta(\mathbf{r} - \mathbf{r}_i) \delta(\mathbf{r}' - \mathbf{r}_j)$ could be the starting operators to calculate next the expectation value of $\hat{\rho}_1(\mathbf{r})$ and $\hat{\rho}_2(\mathbf{r}, \mathbf{r}')$ onto the state $\Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots \mathbf{r}_N)$. The form (6.1) identifies the single-particle density as the probability of finding a particle at position \mathbf{r} , whatever the position of the others might be, so that a factor of N is introduced in front of the integral. The form (6.2) identifies the pair density as the probability of finding one particle at \mathbf{r} and a second one at \mathbf{r}' , independently of the position of the other pairs, so that the number of pairs $N(N-1)/2$ appears in front of the integral.

These definitions become especially transparent if the simplest case is analyzed first. In the following, the case of only two either fermionic or bosonic particles is first discussed, going back to the general case afterward.

6.2.1 Case with two particles

Consider two identical particles with wavefunction $\phi_i(q) = \psi_i(\mathbf{r})\alpha_i(s)$ built up from the single-particle wavefunctions including the orthogonal orbitals $\psi_i(\mathbf{r})$ and the spin part $\alpha_i(s)$. The pair wavefunction is to be antisymmetric in the case of fermions and symmetric in the case of bosons, so that

$$\Phi(q_1, q_2) = \frac{1}{\sqrt{2}} [\phi_1(q_1)\phi_2(q_2) \mp \phi_1(q_2)\phi_2(q_1)], \quad (6.3)$$

where the upper (lower) sign refer to fermion (boson) particles. Here, the factor $1/\sqrt{2}$ accounts for the wavefunction normalization. The probability density $|\Phi(q_1, q_2)|^2$ is in the two cases:

$$\begin{aligned} |\Phi(q_1, q_2)|^2 &= \frac{1}{2} \left[|\phi_1(q_1)|^2 |\phi_2(q_2)|^2 + |\phi_1(q_2)|^2 |\phi_2(q_1)|^2 \right. \\ &\quad \mp \phi_1(q_1)\phi_2(q_2)\phi_1^*(q_2)\phi_2^*(q_1) \\ &\quad \left. \mp \phi_1(q_2)\phi_2(q_1)\phi_1^*(q_1)\phi_2^*(q_2) \right], \end{aligned} \quad (6.4)$$

where again the minus signs refer to fermions and the plus to bosons. When the spin parts of the wavefunctions are equal, the integration over the spin variables defines the probability density of finding one particle at \mathbf{r}_1 and the other at \mathbf{r}_2 :

$$\begin{aligned} |\Psi(\mathbf{r}_1, \mathbf{r}_2)|^2 &= \frac{1}{2} \left[|\psi_1(\mathbf{r}_1)|^2 |\psi_2(\mathbf{r}_2)|^2 + |\psi_1(\mathbf{r}_2)|^2 |\psi_2(\mathbf{r}_1)|^2 \right. \\ &\quad \mp \psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2)\psi_1^*(\mathbf{r}_2)\psi_2^*(\mathbf{r}_1) \\ &\quad \left. \mp \psi_1(\mathbf{r}_2)\psi_2(\mathbf{r}_1)\psi_1^*(\mathbf{r}_1)\psi_2^*(\mathbf{r}_2) \right]. \end{aligned} \quad (6.5)$$

Therefore, (6.1) leads to

$$n(\mathbf{r}) = 2 \int |\Psi(\mathbf{r}, \mathbf{r}_2)|^2 d\mathbf{r}_2 = |\psi_1(\mathbf{r})|^2 + |\psi_2(\mathbf{r})|^2, \quad (6.6)$$

and the pair function $n_2(\mathbf{r}, \mathbf{r}')$ coincides with $|\Psi(\mathbf{r}, \mathbf{r}')|^2$.

Notice the different terms in (6.5): the first line represents the probability of finding the two particles in either one of the two states, while the interference-like terms in the second two lines express the correlations between the two particles. The latter have opposite signs for fermions and bosons. In fact, if $\mathbf{r}_1 = \mathbf{r}_2 = \mathbf{r}$,

one has $|\Psi(\mathbf{r}, \mathbf{r})|^2 = 0$ in the case of fermions and $|\Psi(\mathbf{r}, \mathbf{r})|^2 = 2|\psi_1(\mathbf{r})|^2|\psi_2(\mathbf{r})|^2$ for bosons. If no symmetrization had to be performed, one would obtain the classical probability $|\Psi(\mathbf{r}, \mathbf{r})|^2 = |\psi_1(\mathbf{r})|^2|\psi_2(\mathbf{r})|^2$. Thus, as expected, two identical fermions with the same spin have zero probability to be found at the same position, whereas for two bosons this probability is twice as much it is the probability for the classical case.

By means of (6.6), relation (6.5) can also be cast in the form

$$\begin{aligned} n_2(\mathbf{r}, \mathbf{r}') = \frac{1}{2} & \left[n(\mathbf{r})n(\mathbf{r}') - |\psi_1(\mathbf{r})|^2|\psi_1(\mathbf{r}')|^2 - |\psi_2(\mathbf{r})|^2|\psi_2(\mathbf{r}')|^2 \right. \\ & \mp \psi_1(\mathbf{r})\psi_2(\mathbf{r}')\psi_1^*(\mathbf{r}')\psi_2^*(\mathbf{r}) \\ & \left. \mp \psi_1(\mathbf{r}')\psi_2(\mathbf{r})\psi_1^*(\mathbf{r})\psi_2^*(\mathbf{r}') \right]. \end{aligned} \quad (6.7)$$

This form evidences that since the $n(\mathbf{r})n(\mathbf{r}')$ term would allow the two particles to occupy the same state, the last two terms on the first line serve to exclude this possibility.

In the case of two electron states with opposite spin, integration of (6.4) over the spin variables leads to a vanishing correlation term for both fermions and bosons. This is the reason why the correlation between particles with the same spin is often referred to as spin or exchange correlation. The simple example above shows that

Concept

In quantum world statistics automatically introduces a correlation due to exchange. To this, a correlation originated by the interaction between the particles is to be added.

The simple example does not contain this latter effect since the wavefunction was chosen to be a (anti)symmetrized product of single-particle wavefunctions. Though very common, this is not necessarily the best suited and reasonable choice, since in many other cases the system stability strictly depends on spatial correlations. One such case is the negative hydrogen H^- ion. Expression (6.7) also suggests that

Concept

Exchange has negligible effects whenever the single-particle wavefunctions are localized and is quite effective when the particle wavefunctions are delocalized, as in crystals and especially in metals.

For example, in the case of two interacting nanocrystals with one electron each, exchange is immaterial.

Quick Questions

Q1. Consider three electrons, two of them having the same spin. Discuss the spin correlation in the antisymmetric wavefunction .

Answer - Since state $\phi_3(q)$ works as a spin wavefunction orthogonal to that of the remaining two states, the Slater determinant of third rank is

$$\Psi(q_1, q_2, q_3) = \frac{1}{\sqrt{3!}} \begin{vmatrix} \phi_1(q_1) & \phi_1(q_2) & \phi_1(q_3) \\ \phi_2(q_1) & \phi_2(q_2) & \phi_2(q_3) \\ \phi_3(q_1) & \phi_3(q_2) & \phi_3(q_3) \end{vmatrix}$$

Expansion of it produces the following six terms:

$$\begin{aligned} \Psi(q_1, q_2, q_3) = \frac{1}{\sqrt{3!}} \{ & \phi_3(q_1) [\phi_1(q_2) \phi_2(q_3) - \phi_2(q_2) \phi_1(q_3)] \\ & - \phi_3(q_2) [\phi_1(q_1) \phi_2(q_3) - \phi_2(q_1) \phi_1(q_3)] \\ & + \phi_3(q_3) [\phi_1(q_1) \phi_2(q_2) - \phi_2(q_1) \phi_1(q_2)] \}. \end{aligned} \quad (6.8)$$

Calculating the product $\Phi^(q_1, q_2, q_3)\Phi(q_1, q_2, q_3)$, with the assumption that states $\phi_1(q)$ and $\phi_2(q)$ share the same spin function, and integrating over the spin variables, one obtains*

$$\begin{aligned} |\Psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)|^2 = \frac{1}{6} \big[& |\psi_3(\mathbf{r}_1)|^2 |\Psi(\mathbf{r}_2, \mathbf{r}_3)|^2 \\ & + |\psi_3(\mathbf{r}_2)|^2 |\Psi(\mathbf{r}_1, \mathbf{r}_3)|^2 + |\psi_3(\mathbf{r}_3)|^2 |\Psi(\mathbf{r}_1, \mathbf{r}_2)|^2 \big]. \end{aligned} \quad (6.9)$$

Here, the term $|\Psi(\mathbf{r}_i, \mathbf{r}_j)|^2$ correspond to the result obtained by substituting in (6.5) \mathbf{r}_1 and \mathbf{r}_2 with the quantities \mathbf{r}_i and \mathbf{r}_j . All the interference terms coming from the squared modulus of $\Psi(q_1, q_2, q_3)$ amount to zero because of the orthogonality of the spin functions. Thus, $\psi_3(\mathbf{r})$ does not correlate with the other two, while the remaining two states are correlated within all possible combinations of configurations.

6.2.2 Case with N particles

The functions $n(\mathbf{r})$ and $n_2(\mathbf{r}, \mathbf{r}')$ represent a very useful tool to determine that the energy of a system of N identical particles in the presence of an external field

$$V_{ext}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \sum_{i=1}^N v_{ext}(\mathbf{r}_i), \quad (6.10)$$

is the sum of single-particle potentials and of pair-interaction potentials:

$$U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \frac{1}{2} \sum_{\substack{i,j \\ (i \neq j)}} V(\mathbf{r}_i, \mathbf{r}_j). \quad (6.11)$$

The wavefunction of a bosonic or fermionic system is intended to satisfy the condition $|\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i, \dots, \mathbf{r}_j, \dots, \mathbf{r}_N)|^2 = |\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_j, \dots, \mathbf{r}_i, \dots, \mathbf{r}_N)|^2$ when two out of the N particles are exchanged.

The different contributions are now being expressed in terms of $n(\mathbf{r})$ and $n_2(\mathbf{r}, \mathbf{r}')$.

Single-particle potentials. The contribution of an external potential gets the single-particle expression

$$\begin{aligned} U_{sp} &= \int V_{ext}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) |\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2 d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N \\ &= \sum_{i=1}^N \int v_{ext}(\mathbf{r}_i) d\mathbf{r}_i \int d\mathbf{r}_1 \dots d\mathbf{r}_{i-1} d\mathbf{r}_{i+1} \cdots d\mathbf{r}_N |\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2 \\ &= \int v_{ext}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r}. \end{aligned} \quad (6.12)$$

In the specific case of an electronic system in the presence of an external field originated by charges $Z_I e$ placed at positions \mathbf{R}_I , such as for the electron-nuclei interaction, (6.12) becomes:

$$U_{en} = -e^2 \sum_{I=1}^M Z_I \int \frac{n(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|} d\mathbf{r}. \quad (6.13)$$

Pair potentials. The contribution of a pair interaction potential can be cast into the form

$$\begin{aligned} U_{pp} &= \frac{1}{2} \sum'_{i,j=1}^N \int V(\mathbf{r}_i, \mathbf{r}_j) |\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2 d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N \\ &= \frac{1}{2} \sum'_{i,j=1}^N \int V(\mathbf{r}_i, \mathbf{r}_j) d\mathbf{r}_i d\mathbf{r}_j \int d\mathbf{r}_1 \dots d\mathbf{r}_{i-1} d\mathbf{r}_{i+1} \dots d\mathbf{r}_{j-1} d\mathbf{r}_{j+1} |\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2 \\ &= \int V(\mathbf{r}, \mathbf{r}') n_2(\mathbf{r}, \mathbf{r}') d\mathbf{r} d\mathbf{r}'. \end{aligned} \quad (6.14)$$

Here, the primes on the sums remind that they are to be executed with $i \neq j$. In the specific case of the electrostatic interaction between pairs of electrons, the latter becomes

$$U_{ee} = e^2 \int \frac{n_2(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'. \quad (6.15)$$

The physical implications of (6.15) are evident when analyzing the neat differences with the electrostatic energy between two charge distributions $en(\mathbf{r})$ and $en(\mathbf{r}')$, that is

$$U_{es} = \frac{e^2}{2} \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}'. \quad (6.16)$$

In fact, $n_2(\mathbf{r}, \mathbf{r}')$ accounts for the probability of finding two electrons at \mathbf{r} and \mathbf{r}' , thereby including their exchange and correlation. Besides, substitution of (6.1) in (6.16) would lead to a term proportional to $N^2/2$ in contrast to $N(N-1)/2$ in (6.2). This completely misses to subtract the term $\propto N$: the latter is the sum of N contributions describing the interaction of each charge with itself, what is named a self-energy term. In the absence of correlations between the electrons, but including the self-energy, one would have $n_2(\mathbf{r}, \mathbf{r}') = n(\mathbf{r})n(\mathbf{r}')/2$. As a consequence, n_2 can be cast in the more expressive form

$$n_2(\mathbf{r}, \mathbf{r}') = \frac{1}{2}n(\mathbf{r})n(\mathbf{r}') + \Delta n_2(\mathbf{r}, \mathbf{r}'), \quad (6.17)$$

where $\Delta n_2(\mathbf{r}, \mathbf{r}')$ accounts for correlations and self-energies.

Similar considerations hold for a generic pair interaction. To the lowest order, one has

$$U_{pp} = \frac{1}{2} \int V(\mathbf{r}, \mathbf{r}')n(\mathbf{r})n(\mathbf{r}')d\mathbf{r}d\mathbf{r}', \quad (6.18)$$

where the crucial role of $\Delta n_2(\mathbf{r}, \mathbf{r}')$ emerges, to calculate the interaction potential energy of the system particles. Useful general properties of the single and pair densities and potentials are illustrated in the next section.

6.2.3 The exchange and correlation hole

The definitions (6.1) and (6.2) for $n(\mathbf{r})$ and $n(\mathbf{r}, \mathbf{r}')$ immediately lead to the

Properties

P1.

$$n(\mathbf{r}) = \frac{2}{N-1} \int n(\mathbf{r}, \mathbf{r}')d\mathbf{r}'. \quad (6.19)$$

The correlation terms embodied in $n(\mathbf{r}, \mathbf{r}')$ are best emphasized after writing

$$n(\mathbf{r}, \mathbf{r}') = \frac{1}{2}n(\mathbf{r})n(\mathbf{r}') [1 + h(\mathbf{r}, \mathbf{r}')]. \quad (6.20)$$

Definition

$g(\mathbf{r}, \mathbf{r}') = h(\mathbf{r}, \mathbf{r}') + 1$ is called pair correlation function.

Concept

The quantity $ng(\mathbf{r}, \mathbf{r}')$ can be viewed as the particle density which would be observed at \mathbf{r}' when sitting on the particle at position \mathbf{r} , or viceversa.

Therefore,

$$\Delta n_2(\mathbf{r}, \mathbf{r}') = \frac{1}{2} n(\mathbf{r}) n(\mathbf{r}') h(\mathbf{r}, \mathbf{r}'). \quad (6.21)$$

The properties of $h(\mathbf{r}, \mathbf{r}')$ reflect those of $\Delta n_2(\mathbf{r}, \mathbf{r}')$ and viceversa. In particular, insertion of (6.20) into (6.19) leads to the relation

$$\frac{N-1}{2} n(\mathbf{r}) = \frac{1}{2} n(\mathbf{r}) \left[N + \int d\mathbf{r}' n(\mathbf{r}') h(\mathbf{r}, \mathbf{r}') \right], \quad (6.22)$$

and consequently

$$\int d\mathbf{r}' n(\mathbf{r}') h(\mathbf{r}, \mathbf{r}') = -1, \quad (6.23)$$

whatever \mathbf{r} might be.

These results suggest to introduce the function

$$n_{xc}(\mathbf{r}, \mathbf{r}') = n(\mathbf{r}') h(\mathbf{r}, \mathbf{r}') = n(\mathbf{r}') [g(\mathbf{r}, \mathbf{r}') - 1] \quad (6.24)$$

such that

$$\int n_{xc}(\mathbf{r}, \mathbf{r}') d\mathbf{r}' = -1. \quad (6.25)$$

Therefore, counting all the particles around the chosen one located at \mathbf{r} and while sitting on top of it, amounts to subtract a unit particle. At large distance $|\mathbf{r} - \mathbf{r}'| \rightarrow \infty$ one has $g(\mathbf{r}, \mathbf{r}') \rightarrow 1$ and $h(\mathbf{r}, \mathbf{r}') \rightarrow 0$: this means that one has unit probability of finding a second particle, or else that the system density profile acquires back its homogeneous shape. Therefore, at large distances the exchange and correlation contributions to energy become negligible, an other way to state that the interactions are efficiently screened.

Concept

The density profile $n_{xc}(\mathbf{r}, \mathbf{r}')$ can be viewed as a hole that is dug in by exchange and correlation processes between the system particles, leading to screening of the interactions.

In the specific case in which $V(\mathbf{r}, \mathbf{r}')$ is the bare Coulomb potential, (6.25) would amount to a unit charge of opposite sign with respect to that of the given particle. Then, the pair potential U_{ee} in (6.15) would get the form

$$U_{ee} = U_{es} + \frac{1}{2}e^2 \int d\mathbf{r}d\mathbf{r}' \frac{n(\mathbf{r})n_{xc}(\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \quad (6.26)$$

where U_{es} is the electrostatic energy defined by (6.16). The form (6.26) has the advantage of operating a clear distinction between classical and quantum terms in the Coulomb interaction. The quantum term in particular is discussed below.

Consider the motion of a given electron with given spin and examine the charge distribution of the remaining $N - 1$ electrons. One would expect that this charge density distribution be depleted as a consequence of two physical mechanisms and related properties, as described below:

Properties

P1. Pauli exclusion principle forbids a second electron with the same spin to sit on top of the first: in this case a Coulomb interaction exists between same spin electrons, while no significant Coulomb interaction would take place between opposite spin particles. The given electron with given spin thus interacts with a positive hole charge carrying that given spin. The interaction manifests of course as a Coulomb-like one, though the origin is to be traced back to statistics. The interaction energy associated to this phenomenon is named exchange-induced interaction energy and profiles the so-called Fermi or exchange hole.

P2. Coulomb repulsion induces a charge depletion around the given electron, acting also on electrons with opposite to the given spin. The interaction energy associated to this phenomenon is named correlation-induced interaction energy and profiles the so-called correlation hole.

P3. In both cases, the (negative) electron interacts with a positive charged hole, so that the corresponding contributions to the total energy are negative.

Exchange and correlation hole in the case of bosons. Bose statistics favors the building up of particles at short distances. From (6.21) it turns out that $\Delta n_2(\mathbf{r},\mathbf{r}')$ depends on $h(\mathbf{r},\mathbf{r}')$, whose value is a consequence of interaction between two particles. Therefore, the nature of the interactions implies either positive or negative values of $\Delta n_2(\mathbf{r},\mathbf{r}')$ as compared with that of non-correlated particles $n(\mathbf{r})n(\mathbf{r}')/2$, depending on whether the interactions are repulsive or attractive at short distances. The compared behavior is sketched in Fig. 6.1, where the results of the calculated pair correlation function $g(r)$ are reported for fermions and bosons interacting via the same type of Coulomb potential at two selected values of the coupling strength. The details of the calculation and the approximations used [5] are unimportant here, the focus being on the qualitative behavior. Differences are seen to emerge at weak coupling and short distances: in both cases the value taken by the pair correlation

function at $r = 0$ drops to zero with increasing the coupling strength, that is from the left to the right panel of the figure. However, while the maximum value of $g(0)$ corresponding to the non-interacting system is $1/2$ due to the Pauli exclusion principle for fermions, the maximum value for bosons is 1.

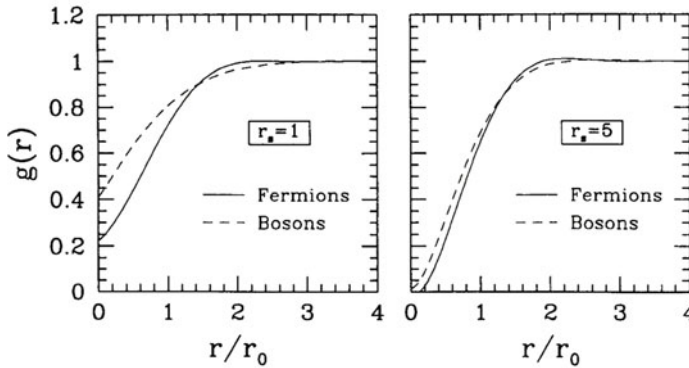


Fig. 6.1 Differential behavior of the exchange and correlation hole for fermions and bosons. The pair correlation function $g(r)$ is displayed as a function of r/r_0 , with r_0 the average interparticle distance, at two different coupling strengths $r_s = r_0/a_b$, with a_b the Bohr radius. Left panel: weaker coupling. Right panel: stronger coupling. In each panel, solid lines represent fermion behavior and dashed lines boson behavior [5], copyright World Scientific Press Co.

The tools developed in the present section are applicable to any system, charged or neutral, boson or fermion-like. In the following, they are being applied to the test case of an electron system within the Hartree-Fock model, for which a number of properties are already known from the analysis of Chap. 2.

6.3 Self-consistent Hartree-Fock potential

In view of the above discussion on the properties of the exchange and correlation hole, the Hartree-Fock model evidently accounts only for the exchange term: the Slater determinant is composed of single-particle orbitals, each of them being the product of an orbital single-particle wavefunction and a spin wavefunction. Therefore, the Hartree-Fock wavefunction completely misses the spatial correlations among the electrons.

The total energy as calculated in Chap. 2 is reminded to be

$$\begin{aligned}
 E(N) = & \sum_{i=1}^N \int \psi_i^*(\mathbf{r}) \left(-\frac{\hbar^2}{2m} \nabla^2 \right) \psi_i(\mathbf{r}) d\mathbf{r} \\
 & - e^2 \sum_{I=1}^M Z_I \int \frac{n(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|} d\mathbf{r} + \frac{e^2}{2} \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \\
 & - \frac{e^2}{2} \sum_{i,j=1}^N \delta_{m_i, m_j} \int \frac{\psi_i^*(\mathbf{r}) \psi_i(\mathbf{r}') \psi_j(\mathbf{r}) \psi_j^*(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}'. \quad (6.27)
 \end{aligned}$$

It is also reminded that application of a variational procedure leads to expressions for the single-particle orbitals, that is

$$-\frac{\hbar^2}{2m}\nabla^2\psi_i(\mathbf{r}) + \{V_{en}(\mathbf{r}) + V_H[n(\mathbf{r})]\}\psi_i(\mathbf{r}) - \sum_{j=1}^N \delta_{m_i, m_j} \int V_F(\mathbf{r}, \mathbf{r}'; j) \psi_j(\mathbf{r}') d\mathbf{r}' = E_i \psi_i(\mathbf{r}). \quad (6.28)$$

Here,

$$V_{en}(\mathbf{r}) = -e^2 \sum_{I=1}^M \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|}, \quad (6.29)$$

$$V_H[n(\mathbf{r})] = e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \quad (6.30)$$

and

$$V_F(\mathbf{r}, \mathbf{r}'; j) = e^2 \frac{\psi_j(\mathbf{r}) \psi_j^*(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (6.31)$$

The potential determining the i -th orbital is self-consistent, that is implicitly contains the knowledge of all the orbitals satisfying (6.28): V_H is in fact a functional of the density $n(\mathbf{r})$. Notice also that at variance with the direct Coulomb term V_H , the exchange term (6.31) has a non-local character, that certainly does not simplify the task. Since the unknown function appears inside the integral in the exchange term, (6.28) can be considered as a set of integral-differential equations. In fact, their solution appears to be a complex task. The contribution with $j = i$ in the exchange term cancels out the piece

$$e^2 \int \frac{|\psi_i(\mathbf{r})|^2}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \psi_i(\mathbf{r}) \quad (6.32)$$

appearing in the third term of (6.28): the self-energy is at once eliminated from both electrostatic and exchange interactions.

In actual calculations, the non-local exchange term is approximated by the local one

$$V_{xc} = -3e^2 \left(\frac{3}{8\pi} n \right)^{1/3}. \quad (6.33)$$

This action is performed by means of the procedure discussed in Sec. 2.5.4, where the exchange term is calculated exactly in the jellium model, using the free particle wave functions, and applying the result also to non homogeneous systems.

The new simplified equation is now to be solved

$$-\frac{\hbar^2}{2m}\nabla^2\psi_i(\mathbf{r}) + \{V_{en}(\mathbf{r}) + V_H[n(\mathbf{r})]\}\psi_i(\mathbf{r}) + V_{xc,loc}[n(\mathbf{r})]\psi_i(\mathbf{r}) = E_i \psi_i(\mathbf{r}). \quad (6.34)$$

Notice that the local approximation for the exchange energy now prevents the cancellation of the exchange and Coulomb self-energy terms.

In Chap. 2 it has been discussed how the eigenvalues E_i represent either the electron ionization or the electron affinity energy, depending on whether the i -th electron is taken away from or added to the N electron system.

When the exchange part $V_{xc,loc}$ is neglected, (6.28) can be written as

$$-\frac{\hbar^2}{2m}\nabla^2\psi_i(\mathbf{r}) + V_{en}(\mathbf{r})\psi_i(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r}') - |\psi_i(\mathbf{r}')|^2}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \psi_i(\mathbf{r}) = E_i \psi_i(\mathbf{r}). \quad (6.35)$$

The Hartree approximation amounts to neglect both the exchange and self-energy terms in (6.35), so that (6.28) becomes

$$-\frac{\hbar^2}{2m}\nabla^2\psi_i(\mathbf{r}) + V_{en}(\mathbf{r})\psi_i(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \psi_i(\mathbf{r}) = E_i \psi_i(\mathbf{r}). \quad (6.36)$$

These are the Hartree equations, characterized by the fact of being a set of formally identical equations for each $\psi_i(\mathbf{r})$ state. This is not the case in (6.35) nor in the Hartree-Fock equation (6.28). There indeed, the direct Coulomb term in (6.35) or the exchange term in (6.28) explicitly depend on state $\psi_i(\mathbf{r})$.

In any event, both Hartree equation (6.36) as well as (6.34) are to be solved in self-consistent manner. To this aim, the following procedure is conveniently adopted:

Procedure

Step 1. An initial trial choice of $\psi_i^{in}(\mathbf{r})$ is first performed.

Step 2. The density $n^{in}(\mathbf{r})$ is calculated by means of $n^{in}(\mathbf{r}) = \sum_i |\psi_i^{in}(\mathbf{r})|^2$.

Step 3. The potential $e^2 \int d\mathbf{r}' n^{in}(\mathbf{r}')/|\mathbf{r} - \mathbf{r}'|$ is then calculated.

Step 4. Equations (6.36) or (6.34) are solved, leading to the determination of $\psi_i^{out}(\mathbf{r})$.

Step 5. The calculation of the density $n^{out}(\mathbf{r})$ is updated with the $\psi_i^{out}(\mathbf{r})$ determined in Step 4, and compared with $n^{in}(\mathbf{r})$. If $|n^{in}(\mathbf{r}) - n^{out}(\mathbf{r})|$ is smaller than a predefined tolerance, the calculation ends with the wavefunctions given by the last iteration step. Otherwise, one sets $n^{in}(\mathbf{r}) = n^{out}(\mathbf{r})$ and iterates back to Step 2.

Rapid convergence within a few iterations can be obtained with suited strategies. In general though, convergence is not necessarily guaranteed.

A different approximation to the Hartree-Fock equation (6.28) is now being considered. As a first step, the exchange term is multiplied and divided by $\psi_i(\mathbf{r}) \psi_i^*(\mathbf{r})$, so that one has:

$$\begin{aligned} & -e^2 \sum_{j=1}^N \int \frac{\psi_i(\mathbf{r}') \psi_j(\mathbf{r}) \psi_j^*(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \frac{\psi_i(\mathbf{r}) \psi_i^*(\mathbf{r})}{\psi_i(\mathbf{r}) \psi_i^*(\mathbf{r})} \delta_{m_i, m_j} \\ & = -e^2 \int \sum_{j=1}^N \frac{n_j^x(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \psi_i(\mathbf{r}), \end{aligned} \quad (6.37)$$

in terms of the exchange density defined by

$$n_i^x(\mathbf{r}, \mathbf{r}') = \sum_{j=1}^N \frac{\psi_i^*(\mathbf{r}) \psi_i(\mathbf{r}') \psi_j(\mathbf{r}) \psi_j^*(\mathbf{r}')}{\psi_i(\mathbf{r}) \psi_i^*(\mathbf{r})} \delta_{m_i, m_j}. \quad (6.38)$$

Equation (6.28) can now be cast in the more compact form

$$\begin{aligned} & -\frac{\hbar^2}{2m} \nabla^2 \psi_i(\mathbf{r}) - e^2 \sum_{I=1}^M \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} \psi_i(\mathbf{r}) \\ & + e^2 \int \frac{n(\mathbf{r}') - n_i^x(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \psi_i(\mathbf{r}) = E_i \psi_i(\mathbf{r}). \end{aligned} \quad (6.39)$$

The dependence on the i -th state of the last term in the left-hand side of (6.39) can be eliminated by averaging $n_i^x(\mathbf{r}, \mathbf{r}')$ over all the states. The averaged exchange density turns out to be

$$n^x(\mathbf{r}, \mathbf{r}') = \frac{\sum_{j=1}^N n_j^x(\mathbf{r}, \mathbf{r}') \psi_j^*(\mathbf{r}) \psi_j(\mathbf{r})}{\sum_{j=1}^N \psi_j^*(\mathbf{r}) \psi_j(\mathbf{r})}. \quad (6.40)$$

Equation (6.39) thus becomes

$$\begin{aligned} & -\frac{\hbar^2}{2m} \nabla^2 \psi_i(\mathbf{r}) - e^2 \sum_{I=1}^M \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} \psi_i(\mathbf{r}) \\ & + e^2 \int \frac{n(\mathbf{r}') - n^x(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \psi_i(\mathbf{r}) = E_i \psi_i(\mathbf{r}). \end{aligned} \quad (6.41)$$

Quick Questions

Q2. Show that (6.38) satisfies

$$\int d\mathbf{r}' n_i^x(\mathbf{r}, \mathbf{r}') = 1$$

whatever \mathbf{r} be. The same occurs for $n^x(\mathbf{r}, \mathbf{r}')$. Discuss the meaning of this result. *Answer - Expression (6.38) straightforwardly leads to the above integral relation. The meaning is understood by noticing that a unit charge is associated to the quantities $n_i^x(\mathbf{r}, \mathbf{r}')$ and $n^x(\mathbf{r}, \mathbf{r}')$, leading to decreasing $n(\mathbf{r})$, as it must be to account for the self-energies.*

6.4 Energy depends on density

The tools developed in Sec. 6.2 and applied in Sec. 6.3 to known approximation schemes, put in bold the dependence of the self-consistent equations on microscopic-defined single-particle and pair densities, embodying a measure of correlations among the particles. In the present section the functional dependence of energy on density is discussed, proceeding through the same approximations as examples with increasing complexity. Yet within these same approximation schemes, the connection with screening and density response theory is illustrated in the next section as one more useful ingredient, preparing to the formal development of the general Density Functional Theory. The material discussed in the present section is referred to fermion Coulomb systems, if not otherwise stated. Generalizations of the results are possible, as e.g. to fermions interacting through different potentials or to interacting bosons: this issue is discussed within the treatment and more systematically considered in Sec. 6.6.

6.4.1 Non interacting particles

Within the jellium model discussed in Chap. 2, the free-particle wavefunctions represented by plane waves have been found to be solutions of the Hartree-Fock set of equations, where the exchange term alone survives: in fact the electrostatic interactions among electrons and jellium and of the electrons among themselves compensate with each other and disappear. The single-particle energy spectrum has been found in the Hartree model to be

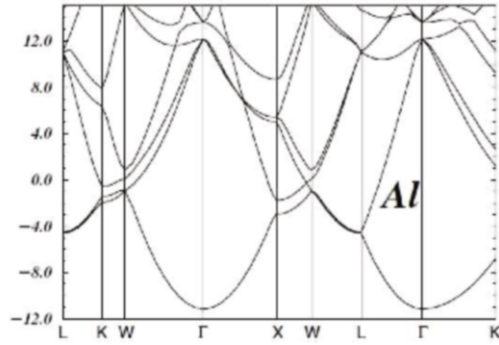
$$E(\mathbf{k}) = \frac{\hbar^2}{2m} k^2. \quad (6.42)$$

The experimental results on metals composed of atoms with *s*- and *p*-electrons in closed internal shells are well enough reproduced by this calculation scheme. A typical example is aluminum, that crystallizes in a face-centered cubic lattice. For example, the band structure as calculated by the pseudopotential method is shown in Fig. 6.2 from [6]. Here, the zero-reference for energy is set to the last occupied state, so that occupied and unoccupied bands have negative and positive energy, respectively. The absence of band gaps between occupied and empty bands typical of metals is also evident. Even more striking is the parabolic-like shape expressed by (6.42) in the first BZ, especially for the first band centered around the Γ point. The pseudopotential theory justifies this interesting result, though a crystal potential is acting, that is non-slowly variable.

The density of states within the jellium model is here reminded as well, that is

$$g(E) = \frac{V}{2\pi^2} \left(\frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \sqrt{E}, \quad (6.43)$$

Fig. 6.2 Hartree approximation scheme for metals. Calculated band structure of aluminum, showing the parabolic free-particle behavior in the first Brillouin zone around the Γ point. The zero reference for energy is set to the last occupied state. Energies are in eV



including a factor of 2 accounting for spin. As a result, the Fermi wavevector k_f , the electron density of the valence (occupied) bands n_0 , and the average energy per particle \bar{E} at zero temperature are all connected to each other:

$$\begin{aligned} k_f &= (3\pi^2 n_0)^{1/3}, \\ \varepsilon_f &= \frac{\hbar^2}{2m} k_f^2 \\ \bar{E} &= \frac{3}{5} \varepsilon_f. \end{aligned} \quad (6.44)$$

A different parameter is conveniently introduced to express the density n_0 of electrons, as follows. The sphere radius r_0 is first defined, that on average contains one electron:

$$\frac{4\pi}{3} r_0^3 n_0 = 1 \rightarrow r_0 = \left(\frac{3}{4\pi n_0} \right)^{1/3}. \quad (6.45)$$

Then, its ratio $r_s = r_0/a_b$ to the Bohr radius $a_b = \hbar^2/me^2$ is introduced, $a_b \simeq 0.529 \times 10^{-8}$ cm. Large $r_s \gg 1$ values correspond to low densities and viceversa. In metals for example, the density n_0 is large in the range of 10^{22} cm $^{-3}$ and r_s gets values between 2 and 6, though most of the metals have r_s in the range between 2 and 3. For aluminum with the band structure in Fig. 6.2, one has e.g. $r_s = 2.07$.

From (6.44) it is easily seen that

$$\begin{aligned} k_f &= \left(\frac{9\pi}{4} \right)^{1/3} \frac{1}{r_0} \rightarrow k_f = \frac{3.63}{r_s} \text{\AA}^{-1}, \\ \varepsilon_f &= \frac{\hbar^2}{2m} \left(\frac{9\pi}{4} \right)^{2/3} \frac{1}{r_0^2} \rightarrow \varepsilon_f = \frac{50.1}{r_s^2} \text{eV}. \end{aligned} \quad (6.46)$$

Thus, the Fermi energy scales as r_s^{-2} . Free electrons at the Fermi level have energies of a few eV. In the same example of aluminum considered above, $r_s = 2.07$ and $\epsilon_f = 11.7$ eV.

The case of bosons. The results above become meaningless in the case of non-interacting bosons [7]. Their statistical behavior would accumulate them all together into the ground state, leading to a macroscopic occupation of the state with zero momentum $p = 0$: at zero temperature all the N particles, if non-interacting, would be there and the kinetic energy would be zero.

6.4.2 The adimensional parameter r_s as a measure of density and coupling strength

Besides being a measure of density, it is quite useful to notice that the adimensional parameter r_s measures the system interaction strength. For example, in Coulomb systems characterized by r^{-1} interactions, $r_s = E_p/E_K$ turns out to be the ratio of the average potential over the average kinetic energy. Therefore, low-density Coulomb systems are characterized by strong interactions, and high-density ones by weak interactions.

To see how this statement works in more general cases and enlighten its connections with fundamental concepts, the starting point is the Virial theorem. While referring to Appendix 6.12 for details, the Virial theorem can be expressed as

$$3PV = 2\langle E_K \rangle - \langle \sum_i \nabla_{\mathbf{r}_i} \Phi \cdot \mathbf{r}_i \rangle, \quad (6.47)$$

in terms of the average kinetic energy E_K , the internal forces $-\nabla_{\mathbf{r}_i} \Phi$ originated by the potential energy Φ of the interacting particles, and the external forces described by pressure P . The latter is the volume derivative of the exact ground-state energy $E_{gs} = E_K + E_p$, that is $P = -dE_{gs}/dV$. The \mathbf{r} and thus density dependence of Φ can be turned into an r_s dependence, and (6.47) provides the ratio E_p/E_K in terms of r_s . Consider the case with $\Phi(\mathbf{r}) \propto \mathbf{r}^\alpha$, that includes a number of situations of interest such as the Coulomb case with $\alpha = -1$. Thus, the Virial theorem gets the form $2E_K - \alpha E_p = 3PV$ or $2E_{gs} - 3PV = (\alpha + 2)E_p$. Assuming that E_{gs} and its potential and kinetic parts be functions of r_s and using the definition of r_s , one has that $Vd/dV = (r_s/3)d/dr_s$ and

$$(\alpha + 2)E_p(r_s) = \frac{1}{r_s} \frac{d}{dr_s} [r_s^2 E_{gs}(r_s)]. \quad (6.48)$$

One example may help to fix the ideas. The electron Fermi gas is characterized by $E_K \approx r_s^{-2}$ and, within the Hartree-Fock theory, one has $E_p \approx r_s^{-1}$: large (small) r_s values correspond to strongly (weakly) interacting regimes where the potential (kinetic) energy dominates. In the extreme dilute regime of large r_s values and strong coupling for example, the electron gas is expected to perform a transition to a new,

lowest-energy quantum state, where the electrons freeze themselves into the positions of a cubic (3D) or hexagonal (2D) lattice, around which they may vibrate: this ground state is named Wigner crystal. Wigner crystals more likely set in under extreme conditions of quantum degeneracy with low temperatures and possibly reduced dimensions, besides the strong correlations as those taking place at very low densities.

One more relevant concept is embodied in (6.48). This is neatly seen after integration, that leads to the expression

$$E_{gs}(r_s) - E_{gs}(r_s = 0) = \frac{1}{r_s^2}(\alpha + 2) \int_0^{r_s} [r_s E_p(r_s)] dr_s. \quad (6.49)$$

The quantity on the right-hand side represents the excess energy that is to be added to the zero-coupling $E_{gs}(r_s = 0)$ energy, in order to get the ground state energy E_{gs} . The key issue here is that

Concept

The excess energy due to exchange and correlations modifies not only the potential energy, but also the kinetic one. This is a purely quantum effect originated by the Heisenberg principle, and occurs whatever the statistics of the particles might be. The adimensional parameter $r_s = r_0/a_b$ measuring the system density can be also considered as a measure of the system coupling strength.

For example, in Hartree-Fock approximation $E_p \approx r_s^{-1}$ and thus the kinetic energy is unchanged with respect to the case of a free-particle gas: and indeed, the Hartree-Fock ground-state wavefunction is the same as that of a perfect Fermi gas.

The above findings can be extended also to bosons. In other cases where the potential has different than \mathbf{r}^α behaviors, the calculation can be adapted to obtain a new differential equation in place of (6.48).

Quick Questions

Q3. Determine the expression of r_s in 2D and 1D.

Answer - Expression (6.45) is substituted by

$$\pi r_0^2 n_0 = 1 \rightarrow r_0^{2D} = \left(\frac{1}{\pi n_0} \right)^{1/2}, \quad (6.50)$$

$$r_0 n_0 = 1 \rightarrow r_0^{1D} = \frac{1}{n_0}. \quad (6.51)$$

The definition $r_s = r_0/a_b$ would not change, and thus $r_s \propto n_0^{-1}$, $n_0^{-1/2}$, and $n_0^{-1/3}$ in 1D, 2D, and 3D respectively.

Q4. In atomic physics, systems of ultracold quantum particles are routinely realized, that are confined in 1D geometries and that interact via dipolar-like potentials $\propto 1/r^3$. Would a dilute such a system with low density correspond to a strongly coupled system?

Answer - No. While kinetic energy would still be scaling as r_s^{-2} , potential energy would be scaling as r_s^{-3} . Therefore, the ratio $E_{pot}/E_k \propto 1/r_s$. At variance with Coulomb system characterized by $\propto 1/r$ thus, low density or large r_s would correspond to weak couplings $E_{pot}/E_k \propto 1/r_s$.

Q5. Consider the Hartree-Fock potential, that scales as $1/r_s$. Is there an excess energy related to the kinetic energy?

Answer - No. From $E_{gs}(r_s) - E_{gs}(r_s = 0) = r_s^{-2}(\alpha + 2) \int_0^{r_s} [r_s E_{pot}(r_s)] dr_s$ and $E_{pot} \propto 1/r_s$, it is immediately seen that no kinetic-like $1/r_s^2$ terms appear in the excess energy.

6.4.3 Interacting particles: direct Hartree contribution

Not much is going to change when interactions are switched on and considered within the Hartree approximation: as already discussed, the many-particle wavefunction corresponding to the ground state is still composed of single-particle orbitals. Thus, the total energy E is the sum of the average kinetic energy per particle over the N particles, that is

$$E = N \frac{3}{5} \varepsilon_f. \quad (6.52)$$

Use of (6.46) yields in Ry:

$$\frac{E}{N} = \frac{30.06}{r_s^2} \text{eV} = \frac{2.21}{r_s^2} \text{Ry}. \quad (6.53)$$

By means of (6.44), this corresponds to the kinetic energy density

$$\frac{E}{V} = \frac{\hbar^2}{2m} \frac{1}{5\pi^2} (3\pi^2 n_0)^{\frac{5}{3}}, \quad (6.54)$$

in terms of the particle density n_0 . One more result is for completeness reminded here from Chap. 2. At zero temperature, the Fermi level coincides with the chemical potential μ . Indeed, μ is the required energy to add a particle to or take it away from the system. Thus,

$$\mu = \frac{\partial E}{\partial N}. \quad (6.55)$$

In the case under consideration, from (6.52) one has that

$$E = \frac{3}{5} N \varepsilon_f = \frac{3}{5} \frac{\hbar^2}{2m} \left(\frac{3\pi^2}{V} \right)^{2/3} N^{5/3}, \quad (6.56)$$

and using (6.55) one immediately sees that $\mu = \varepsilon_f$ at $T = 0$.

Quick Questions

Q6. Calculate the compressibility coefficient B_0 from (6.56) and discuss its meaning for fermions and bosons.

Answer - Since $B_0 = V (\partial^2 E / \partial V^2)_T$, one has

$$B_0 = \frac{\hbar^2}{m} \frac{(3\pi^2)^{2/3}}{3} \left(\frac{N}{V} \right)^{5/3}.$$

The compressibility coefficient increases with increasing the number of particles at given volume and with decreasing volume at given number of particles. For a system of free electrons, the compressibility evidently originates from the Pauli exclusion principle that prevents two fermions with the same spin to occupy the same position. Thus, even if characterized only by its kinetic energy and no interactions, the system cannot be squeezed beyond a threshold, meaning that the system has a positive and finite compressibility. Free bosons can instead be squeezed, as many of them as one wishes. This leads to infinite compressibility.

6.4.4 Interacting particles: exchange Fock contribution

The exchange Fock contribution is now to be added. In Chap. 2, it has been found that within the Hartree-Fock scheme free-particle wavefunctions are solutions of the equations (6.28), leading to the following exchange contribution to the single-particle energy spectrum

$$\begin{aligned} E(k) &= \frac{\hbar^2 k^2}{2m} - \frac{2e^2 k_f}{\pi} g\left(\frac{k}{k_f}\right), \\ g(x) &= \frac{1}{2} + \frac{1-x^2}{4x} \ln \left| \frac{1+x}{1-x} \right|. \end{aligned} \quad (6.57)$$

Expression (6.57) is conveniently cast in terms of r_s as

$$E(k) = \left[\left(\frac{9\pi}{4} \right)^{\frac{2}{3}} \frac{1}{r_s^2} \left(\frac{k}{k_f} \right)^2 - \frac{4}{\pi} \left(\frac{9\pi}{4} \right)^{1/3} \frac{1}{r_s} g\left(\frac{k}{k_f}\right) \right] \text{Ry}. \quad (6.58)$$

Concept

Inclusion of the exchange term works to keep the electrons far apart because of the Pauli exclusion principle and therefore lowers the energy. This effect is expected to become progressively negligible while the density lowers.

This is clearly evident in the prediction (6.58). Fig. 6.3 shows the energy bands determined from (6.58), as functions of k/k_f and for two different values of r_s [6]. It is seen that the difference between the energy calculated with and without exchange contribution fades away with increasing r_s at any given k/k_f , and with increasing k/k_f towards the free-particle regime at any given r_s .

One would at first expect that the inclusion of exchange energy improve the agreement in comparison with experimental data. This is not necessarily the case, as discussed below. Defining ΔE as the energy interval corresponding to occupied states, one has

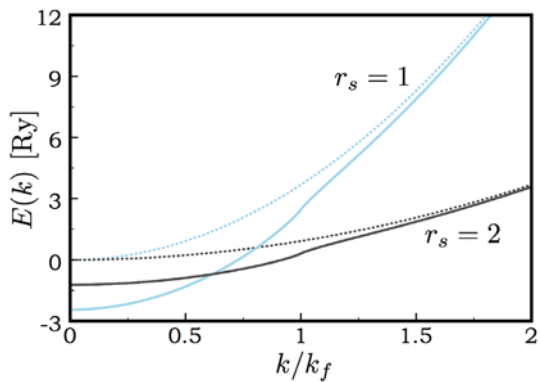
$$\Delta E \equiv E(k_f) - E(0) = \left[\left(\frac{9\pi}{4} \right)^{2/3} \frac{1}{r_s^2} \right] \text{Ry}, \quad (6.59)$$

without the exchange contribution and

$$\Delta E_{HF} \equiv E(k_f) - E(0) = \left[\left(\frac{9\pi}{4} \right)^{2/3} \frac{1}{r_s^2} + \frac{2}{\pi} \left(\frac{9\pi}{4} \right)^{1/3} \frac{1}{r_s} \right] \text{Ry} \quad (6.60)$$

with it. For example, in sodium $r_s = 3.93$ so that from (6.59) one calculates $\Delta E = 3.24$ eV while from (6.60) $\Delta E_{HF} = 7.47$ eV. The experimental data is instead $\Delta E_{\text{exp}} = 2.65$ eV. One therefore sees that in both cases the agreement is not good and actually worsens when exchange is included. This is not just a sodium issue, it is instead common to metals. For example, for aluminum corresponding to the band structure in Fig. 6.2, one has $\Delta E = 11.69$ eV and $\Delta E_{HF} = 19.72$ eV, the experimental value being $\Delta E_{\text{exp}} = 10.06$ eV. Hartree-Fock theory is therefore found to be bad working for metals. Reverting back to the approximations operated so far,

Fig. 6.3 Exchange contributions to energy. Energy band calculated from (6.58) as a function of k/k_f at two selected values of $r_s = 1, 2$. Solid lines: calculation with exchange. Dashed lines: calculation without exchange. The contribution of exchange fades away with increasing r_s , that is after lowering the density



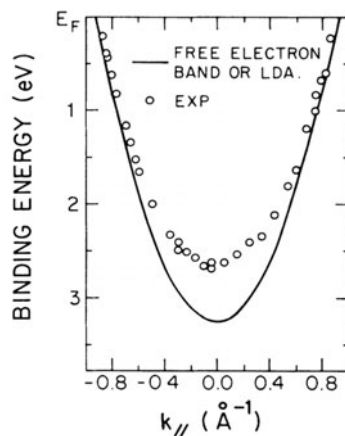
once exchange has been included only spatial correlation effects among electrons are still left completely over. One could therefore expect that

Concept

In metals spatial correlation effects are relevant and are to be treated to some extent together with the spin correlations.

Considering that spatial correlations are long-ranged, one expects as well that consideration of the k dependence be relevant, as it occurs as well for the exchange ones in (6.57). Fig. 6.4 enlightens this point. In the figure, experimental photoemission data are displayed for sodium, that are resolved with respect to the emission angle, thereby providing a measurement of the energy of the occupied band as a function of k wavevector [8].

Fig. 6.4 Relevance of spatial correlations. Angle-resolved photoemission data measuring the occupied band energy of sodium as a function of k . Symbols: experimental data. Solid line: prediction (6.42) for free electrons. The relevance of correlations maximizes in the vicinity of BZ center, that corresponds to long distances in real space. Reprinted with permission from [8]. Copyright (1988) by the American Physical Society



As compared with the free-electron band (6.42) represented by the solid line in Fig. 6.4, the maximum difference is observed at $k = 0$. This suggests that correlation effects be especially relevant in the vicinity of BZ center, in fact corresponding in real space to the long distances characterizing the range of Coulomb interactions. The issue of introducing spatial exchange-correlation effects via their density dependence is the last missing piece in puzzle (6.27). This is introduced next in Sec. 6.5.2, though within yet approximated screening models, as those developed in Chap. 5, and then generalized within the Density Functional Theory.

Quick Questions

- Q7.** Consider either a very dilute system with $r_s \rightarrow \infty$ or a very dense one with $r_s \rightarrow 0$. Discuss the relevance of the exchange term in lowering the energy.

Answer - Consider the second term in (6.58). If $r_s \rightarrow 0$, then $k_f \rightarrow \infty$ and $k/k_f \rightarrow 0$. Therefore, $g(k/k_f) = 1$ and the absolute value of the exchange term is enhanced, scaling as r_s^{-1} . If $r_s \rightarrow \infty$, then $k_f \rightarrow 0$ and $k/k_f \rightarrow \infty$. Therefore $g(k/k_f) = 0$ and the exchange contribution tends to vanish.

6.5 Introducing spatial inhomogeneity via local density

Qualitative changes are met as soon as the jellium scheme is abandoned, namely the ionic potential is not any longer considered as a positive charge uniformly distributed over the crystal: this affects the electron density that acquires a space dependent profile $n(\mathbf{r})$. In this section, spatial exchange-correlation effects are introduced via their density dependence and within approximated models, reminding the generalization to Sec. 6.6.

6.5.1 Kinetic and interaction energy: Thomas-Fermi approximation

The kinetic energy is considered first. In the presence of a space dependent $n(\mathbf{r})$, the link between the kinetic energy density and the particle density is not any longer expressed by (6.54). An updated expression can be obtained under the assumption that the degree of density inhomogeneity $\delta n(\mathbf{r}) = n(\mathbf{r}) - n_0$ is a slowly-varying function of position, so that the potential can be considered to be uniform within a small volume portion of the system, while that volume is at the same time large enough to contain a large number of particles. Within each one of these portions, the relations found for the free particles are still valid to a good extent. The kinetic energy term can therefore be approximated by the free-particle expression (6.54) of the homogeneous system, calculated at the local density $n(\mathbf{r})$ of the weakly-inhomogeneous one.

The new situation can be analyzed to first approximation by reverting back to (6.27). In order to proceed gradually, the different pieces are added step by step, introducing one complication at a time. In particular, the Fock exchange part of the total energy is left on a side, and is considered in Sec. 6.4.4.

Collecting all the above considerations together, (6.36) can be approximated by the following expression:

$$\begin{aligned}
 E[n] = & \frac{\hbar^2}{2m} \frac{(3\pi^2)^{5/3}}{5\pi^2} \int n(\mathbf{r})^{5/3} d\mathbf{r} - e^2 \sum_{I=1}^M Z_I \int \frac{n(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|} d\mathbf{r} \\
 & + \frac{e^2}{2} \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}'.
 \end{aligned} \tag{6.61}$$

The key issue in this expression, quite interesting to the development of knowledge in the present Chapter, is that

Concept

The total energy is expressed as a functional of the electron density alone, the kinetic energy term being given by the free-particle expression of the homogeneous gas, calculated at local density $n(\mathbf{r})$ of the weakly-inhomogeneous system.

The set of conditions used to obtain (6.61) includes the neglect of exchange effects and space correlations, and the approximation of the kinetic energy with that of the free electrons in the homogeneous system but calculated at the local system density. The electron-nuclei interactions is instead exact. The set of conditions expressing (6.61) corresponds to the Thomas-Fermi approximation.

In spite of the underlying approximations, the great advantage of using (6.61) is by far evident, when calculating the structure of many-particle systems such as atoms, molecules, or solids. As seen in Sec. 6.2, general schemes based on the knowledge of the wavefunctions require the solution of a complex Schrödinger equation of $3N$ variables, as many as the electron degrees of freedom are. If instead the primary quantity can be identified with the density $n(\mathbf{r})$, the number of degrees of freedom is reduced to only the three coordinates. As seen from both the theoretical and the practical computational perspectives, this finding has marked a pivotal cornerstone in the treatment of interacting many-particle systems and opened the way towards the Density Functional Theory.

Quick Questions

Q8. To interpret some given experimental data, you decide and try to add

a gradient expansion of the form $\nabla n(\mathbf{r})$ into the kinetic energy term. Using dimensional considerations, elaborate the form that this term should have.

Answer. Not considering the factor \hbar^2/m , the quantity $[n(\mathbf{r})]^{-1/3}$ is a local characteristic length of the system. Therefore from dimensional considerations, the new term should be of the form $\propto [n(\mathbf{r})]^{-2/3} [\nabla n(\mathbf{r})^{5/6}]^2$ that is $\propto [\nabla n(\mathbf{r})]^2 / n(\mathbf{r})$.

6.5.2 Exchange-correlation and screening effects

The analysis of the exchange and correlation leads to a picture where interactions affect the probability of finding a particle at a given distance from a reference one and thus determine local changes in the density profile. The pair correlation function

$g(\mathbf{r}, \mathbf{r}')$ introduced in Chap. 1 and here discussed is a useful tool to define the exchange and correlation hole and describe such processes. In essence, while a system with non-interacting particles is characterized by homogeneous density profiles, interacting systems can be viewed as if they were composed of particles carrying on their own exchange and correlation hole. The hole size is related to the range after which the original bare interactions are effectively screened. When these effective particles are considered, the homogeneity characterizing the non-interacting system is somehow restored. In fact, one could think to describe the inhomogeneous system of interacting particles in terms of a homogeneous non-interacting system composed of such effective particles. This is indeed the basic idea. In this section the link between the exchange-correlation hole picture and screening effects is illustrated within specific approximations that are already at hand. This is generalized in Sec. 6.6.

Screening effects have been introduced in Chap. 5 as a useful concept to describe the optical properties of solids, and the connection between screening and the \mathbf{q} and ω dependent dielectric function has been explored within selected approximations. In particular, it has been seen that the static $\omega = 0$ limit of the dielectric function describes the screening of electrostatic interactions between the particles. The main results obtained there, are summarized for convenience in the following. First, the longitudinal dielectric function in the case of external potentials that are slowly-varying in time is defined by

$$\epsilon_L(\mathbf{q}) = \frac{\phi_{\text{ext}}(\mathbf{q})}{\phi(\mathbf{q})}, \quad (6.62)$$

in terms of the total ϕ and external ϕ_{ext} potentials.

The charge fluctuation $\rho_{\text{ind}}(\mathbf{q}) = \rho(\mathbf{q}) - \rho_0(\mathbf{q}) = -e[n(\mathbf{q}) - n_0(\mathbf{q})]$ induced by the action of the external field is determined by the system response function $\chi(\mathbf{q})$ by means of

$$e[n(\mathbf{q}) - n_0(\mathbf{q})] = e n_{\text{ind}}(\mathbf{q}) = \chi(\mathbf{q}) \phi_{\text{ext}}(\mathbf{q}). \quad (6.63)$$

The density response function $\chi(\mathbf{q})$ and the longitudinal dielectric function $\epsilon_L(\mathbf{q})$ are related via the Poisson equation $\phi(\mathbf{q}) - \phi_{\text{ext}}(\mathbf{q}) = -v_{\mathbf{q}} e n_{\text{ind}}(\mathbf{q})$ and (6.62) in terms of the Fourier transform $v_{\mathbf{q}} = 4\pi/q^2$ of the bare Coulomb potential, that is $1/\epsilon_L(\mathbf{q}) - 1 = -v_{\mathbf{q}} \chi(\mathbf{q})$. Finally, in real space the induced density fluctuation is

$$\rho_{\text{ind}}(\mathbf{r}) = e \int \frac{d\mathbf{q}}{(2\pi)^3} e^{i\mathbf{q} \cdot \mathbf{r}} \left[\frac{1}{\epsilon(\mathbf{q})} - 1 \right], \quad (6.64)$$

as found in (5.112).

6.5.3 Thomas-Fermi linear screening

Consider the case of a free atom with ionic charge eZ or an impurity with equivalent charge placed in a jellium at position $r = 0$. The calculation of $\varepsilon(\mathbf{q})$ within the Thomas-Fermi model (6.61) is performed by using the formalism above, in order to determine the potential energy $V(\mathbf{r})$ that solves the equation

$$\nabla^2 V(r) = 4\pi e^2 [(n(r) - n_0)]. \quad (6.65)$$

Since the point-like charge eZ is placed at the origin, (6.65) is to be solved with the boundary condition

$$\lim_{r \rightarrow 0} rV(r) = -Ze^2. \quad (6.66)$$

A second equation is now needed that links the potential energy V to the electron density n . This can be found starting from the total energy functional in the Thomas-Fermi approximation. To this aim, the functional can be rewritten placing the charge eZ at the origin, so that one has [9]

$$E[n(\mathbf{r})] = \frac{\hbar^2}{2m} \frac{(3\pi^2)^{5/3}}{5\pi^2} \int n(\mathbf{r})^{5/3} d\mathbf{r} - \int \frac{Ze^2 n(\mathbf{r})}{|\mathbf{r}|} d\mathbf{r} + \frac{1}{2} \iint \frac{e^2 n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'. \quad (6.67)$$

In the presence of the charge Ze electrons redistribute themselves so to minimize the total energy (6.67). To fall into the minimum, the total number of electrons is to be conserved. This leads to the conditioned minimum equation

$$\delta(E - \mu N) = 0, \quad (6.68)$$

where the total number of particles

$$N = \int n(\mathbf{r}) d\mathbf{r} \quad (6.69)$$

fixes the chemical potential μ defined by (6.55). Using the rules to perform variational calculations introduced in Appendix 2.10 of Chap. 2, the stationarity condition for the functional $E[n(\mathbf{r})] - \mu \int n(\mathbf{r}) d\mathbf{r}$ can be calculated. This yields the equation

$$\frac{\hbar^2}{2m} (3\pi^2 n(\mathbf{r}))^{2/3} - \frac{Ze^2}{|\mathbf{r}|} + e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' - \mu = 0. \quad (6.70)$$

Setting the total potential energy to be

$$V(\mathbf{r}) = -\frac{Ze^2}{|\mathbf{r}|} + e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \quad (6.71)$$

(6.70) becomes

$$\frac{\hbar^2}{2m} [3\pi^2 n(\mathbf{r})]^{2/3} + V(\mathbf{r}) - \mu = 0, \quad (6.72)$$

where the following relation can be obtained linking potential V and density $n(\mathbf{r})$, that is:

$$n(\mathbf{r}) = \left(\frac{2m}{\hbar^2} \right)^{3/2} \frac{1}{3\pi^2} [\mu - V(\mathbf{r})]^{3/2}. \quad (6.73)$$

This relationship between density and potential is non linear. Consistently with the approach in Sec. 6.5.2, linearization is to be performed. To this aim, the chemical potential μ is conveniently written as

$$\mu = \varepsilon_f + \delta\mu, \quad (6.74)$$

that is in terms of the equilibrium chemical potential ε_f at zero temperature plus a fluctuation $\delta\mu$ due to the external perturbation. Since one expects that $V \ll \varepsilon_f$, linearization of (6.73) is better performed with respect to the ratio V/ε_f . This yields

$$n(\mathbf{r}) = \left(\frac{2m}{\hbar^2} \right)^{3/2} \frac{1}{3\pi^2} \varepsilon_f^{3/2} \left[1 + \frac{3}{2} \frac{\delta\mu - V(\mathbf{r})}{\varepsilon_f} \right] = \frac{k_f^3}{3\pi^2} \left[1 + \frac{3}{2} \frac{\delta\mu - V(\mathbf{r})}{\varepsilon_f} \right]. \quad (6.75)$$

At this point the equilibrium solution can be subtracted, since in (6.65) the fluctuation $n(\mathbf{r}) - n_0$ appears indeed. Therefore, subtracting from (6.75) the density $n_0 = k_f^3/(3\pi^2)$ one obtains:

$$\delta n(\mathbf{r}) = n(\mathbf{r}) - n_0 = \frac{mk_f}{\pi^2 \hbar^2} [\delta\mu - V(\mathbf{r})]. \quad (6.76)$$

Substituting the latter into (6.65), the following equation is obtained for the potential energy V :

$$\begin{aligned} \nabla^2 V(\mathbf{r}) &= k_0^2 [V(\mathbf{r}) - \delta\mu], \\ k_0^2 &= \frac{4}{a_b \pi} k_f. \end{aligned} \quad (6.77)$$

Expression (6.77) represents the Thomas-Fermi linearized equation for the potential V .

The general solution has spherical symmetry and is given by

$$V(r) = -\frac{1}{r} \left(A e^{-k_0 r} + B e^{k_0 r} \right) + \delta\mu. \quad (6.78)$$

Boundary conditions are to be used to finalize the solution of (6.78) besides that of point-like charge at the origin.

These additional boundary conditions depend on the specific nature of the material. Different boundary conditions lead to different solutions. In a covalent insulator for example, the electrons are mainly localized along the directions of the atomic bonds, whereas in a conductor no localization occurs. Appendix 6.13 discusses the particular solution that is obtained with boundary conditions appropriate to insulators, where the density fluctuation $n(r) - n_0$ can be considered to be limited to a

Table 6.1 Screening in metals and insulators. Static dielectric constant ϵ_0 , screening wavevector k_0 and radius R_s are listed for diamond, silicon and aluminum. For comparison, characterizing material parameters are also listed, that are lattice parameter a , density n_0 and Fermi wavevector k_f , gap energy E_g , and distance between first-neighbor atoms d .

| Quantity | Diamond | Silicon | Aluminum |
|-----------------------------|-----------------------|-----------------------|-----------------------|
| a (Å) | 3.57 | 5.43 | 4.05 |
| n_0 (cm ⁻³) | 7.03×10^{23} | 1.99×10^{23} | 1.81×10^{23} |
| E_g (eV) | 5.5 | 1.1 | 0 |
| ϵ_0 | 5.7 | 11.4 | ∞ |
| k_f (a.u. ⁻¹) | 1.45 | 0.96 | 0.92 |
| k_0 (a.u. ⁻¹) | 1.36 | 1.10 | 1.08 |
| R_s (Å) | 1.46 | 2.24 | ∞ |
| d (Å) | 1.54 | 2.35 | 3.51 |

sphere with radius R_s centered on the charge eZ . The detailed calculation leads to the result that

Concept

The static dielectric function ϵ_0 measures the number N_i of induced charges contained within the screening sphere with radius R_s via

$$N_i = Z \left(1 - \epsilon_0^{-1} \right),$$

(6.79)

with ϵ_0 related to the boundary conditions through $\epsilon_0 = \sinh(k_0 R_s) / (k_0 R_s)$. In an insulator $N_i < Z$, whereas in metal $N_i = Z$, as shown in Chap. 5, implying $\epsilon_0 \rightarrow \infty$ in (6.79) and $R_s \rightarrow \infty$.

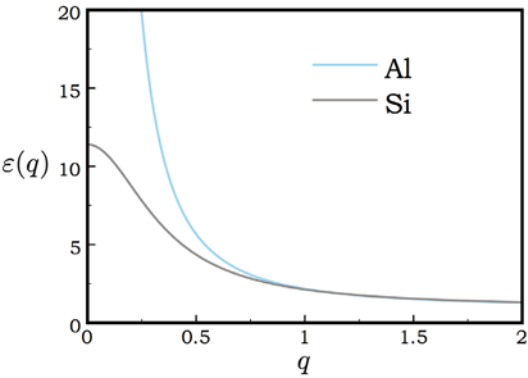


Fig. 6.5 Screening in metals and insulators. The Thomas-Fermi approximation to the dielectric function as a function of q is displayed for silicon and aluminum [10]

Table 6.1 lists the experimental data for lattice parameter a , density n_0 , gap energy E_g , and static dielectric constant ϵ_0 , along with k_0 , R_s , and the distance d between neighbor atoms for diamond, silicon, and aluminum. A few interesting connections are observed. First, the link between ϵ_0 and gap E_g , that has already been remarked in Chap. 5: the smaller is E_g , the larger is ϵ_0 . Second, R_s in diamond and silicon are comparable with the distance d between first-neighbor atoms in the lattice, justifying a posteriori the initial assumption. Last, increasing values of the electronic densities correspond to increasingly larger static dielectric constant, in fact $\epsilon_0 \rightarrow \infty$. **Fig. 6.5** displays the diverging vs. finite behavior of $\epsilon(q \rightarrow 0)$ for aluminum and silicon, respectively.

The procedure used to determine the screening behavior within the Thomas-Fermi model can be summarized as follows:

Procedure

Step 1. Start from the energy functional (6.67) with condition (6.69) fixing the total number of particles.

Step 2. Set stationarity condition $\delta(E - \mu N) = 0$ to obtain $n(r)$ as a function of μ and $V(r)$ as in (6.70).

Step 3. In order to push the analytical calculation as far as possible and whenever applies, linearize (6.73) and calculate $\delta n(r)$ as in (6.76) after subtracting the equilibrium solution, and $V(r)$ with suited boundary conditions.

Step 4. To get additional information, the total number of induced charges N_i within the screening sphere can be calculated as in (6.79) using suited boundary conditions.

Step 5. Calculate $\rho_{\text{ind}}(\mathbf{r})$ from the knowledge of $V(r)$ via (6.76) and (6.78) or $\rho_{\text{ind}}(\mathbf{q})$ via (6.64) and the knowledge of $\epsilon(q)$.

Steps of the procedure outlined above are now being applied in a number of examples.

Examples

Total energy of a neutral atom and of selected molecules. Multiplying (6.70) times $n(\mathbf{r})$ and integrating over \mathbf{r} one obtains

$$\begin{aligned}\mu N &= \frac{5}{3}E_c + U_{en} + 2U_{es}, \\ U_{en} &= - \int d\mathbf{r} \frac{Ze^2}{r} n(\mathbf{r}),\end{aligned}$$

$$U_{es} = \frac{1}{2}e^2 \int d\mathbf{r}d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \quad (6.80)$$

where E_c is the kinetic energy. Equation (6.73) implies that $\mu = 0$ when $n(\mathbf{r}) = 0$ in the limit $r \rightarrow \infty$. Thus, for neutral atoms $\mu = 0$ and (6.80) yields

$$\begin{aligned} -\frac{5}{3}E_c &= U + U_{es} \\ U &= U_{en} + U_{es}. \end{aligned} \quad (6.81)$$

The total energy amounts to $E = E_c + U$: this results from (6.67) and the consideration that the Thomas-Fermi model is semiclassical. Appendix 6.12 recalls the proof of the Virial theorem for quantum systems in the absence of external forces due to pressure. Under this condition, the theorem reads

$$2E_c + U = 0. \quad (6.82)$$

This is valid for the Thomas-Fermi model too, so that

$$E_c = -E. \quad (6.83)$$

This leads to the calculation of U_{en} and U_{es} as functions of E_c or E . One obtains that

$$\begin{aligned} U_{en} &= -\frac{7}{3}E_c = \frac{7}{3}E, \\ U_{es} &= \frac{1}{3}E_c = -\frac{1}{3}E. \end{aligned} \quad (6.84)$$

Using accurate numerical self-consistent calculations, it has been demonstrated that scaling relations exist between the different contributions to the total energy E and tested to a good approximations for both atoms and ions. It can also be demonstrated that the relation

$$E = -0.7687 \cdot Z^{7/3} \frac{e^2}{a_b}$$

holds. The dependence of E on Z is consistent with a system composed of Z electrons interacting with a nucleus with charge Z , and placed at an average distance proportional to $Z^{-1/3}$. The proportionality constant amounts to about 20.91 eV, whereas its empirical value is about 16 eV. It must be recalled that the Thomas-Fermi model is semiclassical, therefore is expected to provide bad results in comparison with experiments whenever the potential $V(r)$ is rapidly varying or whether not enough particles are in the single small volume. The former condition is met in the vicinity of the origin and the latter at large $r \rightarrow \infty$ distances. Irrespective of these limits of validity, self-consistent calculations for atoms provide a good testing of relations (6.84). The Thomas-Fermi model does not admit the formation of molecules. However, if molecules of the type CH_4 , GeH_4 , ... are considered and the hydrogen charges are substituted by spherical surface charge distributions, solutions exist. Considering that the total energy is now

$$E = E_c + U_{en} + U_{es} + U_{n-n}. \quad (6.85)$$

Adding to U_{en} and U_{ee} an interaction term between the nuclei and using the Virial theorem, the following scaling laws are found:

$$\begin{aligned} \frac{U_{es} - U_{n-n}}{E_c} &= \frac{1}{3}, \\ \frac{U_{en} + 2U_{n-n}}{E_c} &= -\frac{7}{3}, \\ \frac{U_{en} + 2U_{es}}{E_c} &= -\frac{5}{3}. \end{aligned} \quad (6.86)$$

The third relation is obtained after summing twice the first to the second equation. Once again, the self-consistent calculations of molecular energies satisfy the scaling laws.

Examples

Relationship between total and single-particle energies. From (6.36) the self-consistent calculation of single-particle orbitals can be performed within the Hartree scheme: multiplying the i -th equation times $\psi_i^*(\mathbf{r})$, integrating and summing over indexes i , one obtains

$$\begin{aligned} E_s &= \sum_i E_i = E_c + \sum_i \int \psi_i^* V \psi_i d\mathbf{r}, \\ V &= V_n + V_e, \\ V_n &= -\frac{Ze^2}{|\mathbf{r}|}, \\ V_e &= \int d\mathbf{r}' \frac{n(\mathbf{r}')e^2}{|\mathbf{r} - \mathbf{r}'|}. \end{aligned} \quad (6.87)$$

This can also be written

$$E_s = \sum_i E_i = E_c + U_{en} + 2U_{es}. \quad (6.88)$$

Using relations (6.84) for neutral atoms, one has

$$E = \frac{3}{2} E_s. \quad (6.89)$$

This result is interesting since it accounts for a quantitative difference between energies E and E_s that, as already shown, are to be different: E_s represents the electrostatic interaction energy between electrons and is counted twice. For a positive ion one finds instead $E = (3/2)(E_s - N\mu)$.

Examples

Thomas-Fermi model for ultracold bosonic atoms. Quantum behavior of bosonic particles is ruled by Bose-Einstein statistics. While referring to Appendix 2.13 of Chap. 2 for details, an example is given below of existing bosonic systems whose ground state can be effectively determined by using the Thomas-Fermi approximation. These are trapped alkali atoms cooled down to a few tens of nK temperatures while remaining in a very dilute gaseous state. Below a critical temperature T_c that is found to be in this nK range, the atoms collectively execute the transition to a different state of matter, that can be viewed as a condensation in momentum state: this is a manifestation of the phenomenon of Bose-Einstein condensation. In original experiments the confinement is provided by magnetic fields and has an accurate parabolic shape as described by a harmonic-oscillator potential. Following the pedagogical theoretical discussion of the original experiments given in [11], consider the wavefunction of a bosonic atom with mass M trapped in a parabolic external potential $V_{\text{ext}}(r) = M\omega_c^2 r^2/2$ with ω_c the cyclotron frequency. The wavefunction is that corresponding to the ground state of a harmonic oscillator, therefore:

$$\phi_0(r) = \frac{1}{\pi^{3/4} a_0^{3/2}} e^{-M\omega_c^2 r^2/2\hbar}, \quad (6.90)$$

where the harmonic oscillator length $a_0 = \sqrt{\hbar/(M\omega_c)}$ has been conveniently introduced. The bosonic atoms love to stack together and collectively undergo Bose-Einstein condensation, leading to a macroscopic occupation of the ground state as described in Appendix 2.13 of Chap. 2.

In fact, atoms interact among themselves: the interaction is accurately enough described by means of a contact pseudopotential $U(r) = 4\pi\hbar^2 a \delta(\mathbf{r})/M$ driven by one single parameter that is the scattering length a , besides the atom mass M . The diluteness condition $na^3 \ll 1$ guarantees the weakly interacting regime. The existence of this repulsive interaction, though weak, ensures the stability of the many boson systems in the ground state against collapse. In the presence of such a contact interaction potential, the equation for the wavefunction $\psi(\mathbf{r})$ gets the form of a non-linear Schrödinger equation, known as Gross-Pitaevskii equation:

$$\left[-\frac{\hbar^2}{2M} \nabla^2 + \frac{1}{2} M \omega_c^2 r^2 + \frac{4\pi\hbar^2 a}{M} |\psi(\mathbf{r})|^2 \right] \psi(\mathbf{r}) = \mu \psi(\mathbf{r}). \quad (6.91)$$

The particle density is given by

$$n_0(r) = N |\psi(\mathbf{r})|^2. \quad (6.92)$$

If the interaction term were neglected, $\psi(\mathbf{r}) = \phi_0(\mathbf{r})$ and the solution reduces to (6.90), as for a harmonic oscillator. To catch the effects of the interaction term,

the kinetic energy term driven by the laplacian in (6.91) can be treated within the Thomas-Fermi approximation. In the confined geometry under consideration, the local density is expected to be almost flat inside the harmonic trap and to get vanishingly small values at the trap boundaries: in the present limit of very large number N of particles, the Thomas-Fermi approximation here amounts to neglect kinetic energy whatsoever. A look at the size of the different energy terms helps understanding this point. If the condensate extends over the spatial size $\approx R$, the kinetic energy scales as $E_K \approx \hbar^2/(2MR^2)$, the confinement energy as $E_C \approx M\omega_c^2 R^2/2$ and the interaction energy as $E_I \approx (4\pi\hbar^2 a/M)(N/R^3)$. Thus, after neglecting E_K to first approximation at equilibrium, R scales according to the relation $R^5 \approx 8\pi\hbar^2 a N / (M^2 \omega_c^2) = a_0^5 (8\pi a N / a_0) \equiv a_0^5 \zeta^5$. Inserting this result into the different energy scales, one has that $E_K \approx \zeta^{-2} \hbar \omega_c / 2$, E_C and $E_I \approx \zeta^2 (\hbar \omega_c / 2)$. In conclusion, E_K/E_C and $E_K/E_I \approx \zeta^{-4}$ that is quite negligible in the large N limit. As a result, to first approximation the kinetic energy can be dropped out. From (6.91), $|\psi(\mathbf{r})|^2$ can therefore be calculated to give

$$|\psi(\mathbf{r})|^2 = \frac{M}{4\pi\hbar^2 a} \left(\mu - \frac{1}{2} M \omega_c^2 r^2 \right). \quad (6.93)$$

The atomic density profile reflects the shape of the external potential, as it is indeed observed in experiments by absorption imaging. Imposing that the total number of particles in the volume be N , provides the expression for the chemical potential $\mu = (\hbar\omega_c/2)\xi^2$, where the parameter $\xi = (15Na/a_0)^{1/5}$ has been conveniently defined. The function describing the way in which the momenta of the atoms in the condensate are distributed, is obtained by Fourier transform of (6.93), obtaining the momentum distribution $f(\mathbf{p}) \approx |J_2(\kappa)/\kappa^2|^2$ where J_2 is the Bessel function of second order and $\kappa^2 = (\xi a_0)^2 p^2$. In the limit $Na/a_0 \gg 1$ of very large number of atoms, the width of the momentum distribution scales as $1/\xi$ times that for a single particle in a harmonic oscillator potential. The momentum distribution is therefore extraordinarily much thinner than that of many independent non-condensed classical particles: the N Bose-condensed particles behave as if they were one single particle with very well defined momentum.

Quick Questions

Q9. Relation (6.73) comes from the minimization of total energy (6.67)

with respect to density $n(\mathbf{r})$. What is the meaning of this operation? .

Answer - In the case under consideration, the total energy is a function of the parameter $n(r)$, so that the equation follows from requiring that the total energy at equilibrium be the least possible.

6.5.4 Exchange energy of the homogeneous electron gas: the Thomas-Fermi-Dirac model

The Thomas-Fermi model misses the exchange part of the exchange and correlation hole. This is being considered as follows. Keeping the main idea of the Thomas-Fermi model, an useful approach is to revert back to the exchange term in the homogeneous system, and then operate for its local-density approximation as done for the kinetic energy term.

The exchange contribution in Hartree-Fock model is given by the non-local potential (6.28), subsequently transformed into the local-like (6.33). If the total electronic energy is considered as in (6.27) and the exchange term is manipulated by means of the same procedure as that used for the potential, the density of exchange energy is

$$\begin{aligned} e_x &= -C_e n^{4/3}, \\ C_e &= \frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} e^2. \end{aligned} \quad (6.94)$$

The functional derivative $\delta e_x / \delta n$ provides the exchange potential with the same dependence on n as in (6.33), the only difference being in the factor $3/2$.

The density dependence of kinetic and exchange energy density can be inferred by means of dimensional analysis. The kinetic energy density is to be proportional to \hbar^2/m , so that

$$\left[\frac{E}{V} \right] = [n(\mathbf{r})]^a \left[\frac{\hbar^2}{m} \right] \rightarrow \frac{ML^2}{T^2 L^3} = \frac{1}{L^{3a}} \frac{M^2 L^4 T^2}{T^4 M} \rightarrow 3a - 4 = 1 \rightarrow a = \frac{5}{3}. \quad (6.95)$$

In the case of exchange energy originated by the Coulomb interaction between the particles, one expect that be proportional to e^2 so that:

$$\left[\frac{E}{V} \right] = [n(\mathbf{r})]^b [e^2] \rightarrow \frac{[E]}{L^3} = \frac{1}{L^{3b}} [E] L \rightarrow 3b - 1 = 3 \rightarrow b = \frac{4}{3}. \quad (6.96)$$

The new functional can thus be cast in the form

$$\begin{aligned} E[n(\mathbf{r})] &= C_k \int [n(\mathbf{r})]^{5/3} d\mathbf{r} + \int n(\mathbf{r}) V_n(\mathbf{r}) d\mathbf{r} \\ &+ \frac{1}{2} e^2 \int d\mathbf{r} d\mathbf{r}' \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} - C_e \int [n(\mathbf{r})]^{4/3} d\mathbf{r}, \end{aligned} \quad (6.97)$$

where $C_k = [\hbar^2/(2m)][(3\pi^2)^{5/3}/(5\pi^2)]$. Imposing the stationarity condition for $E[n(\mathbf{r})]$ with respect to variations of the density $n(\mathbf{r})$ and the constraint (6.69), as already operated for (6.67) one obtains the Thomas-Fermi-Dirac equation:

$$\mu = \frac{5}{3} C_k [n(\mathbf{r})]^{2/3} + V_n(r) + V_e(r) - \frac{4}{3} C_e [n(\mathbf{r})]^{1/3}. \quad (6.98)$$

Expression (6.98) is a second order algebraic equation in $[n(\mathbf{r})]^{1/3}$. The condition $\mu > V(\mathbf{r}) = V_n(\mathbf{r}) + V_e(\mathbf{r})$ guarantees that $n(\mathbf{r})$ be positive. With it, only one of the two solutions is meaningful and the following relation between the density and the potential is obtained

$$n(\mathbf{r}) = \frac{1}{\left(\frac{5}{3}C_k\right)^{3/2}} \left(D + \sqrt{D^2 + \mu - V(\mathbf{r})} \right)^3, \\ D^2 = \frac{\left(\frac{2}{3}C_e\right)^2}{\frac{5}{3}C_k} = \frac{2me^4}{h^2}. \quad (6.99)$$

Inclusion of exchange can thus be viewed as an application of the Thomas-Fermi model to the single-particle Schrödinger equation, though with a potential that adds to the Hartree part $V_n + V_e$ the exchange term

$$V_{eff}(\mathbf{r}) = V_n + V_e + V_x. \quad (6.100)$$

Here,

$$V_x = -e^2 \left(\frac{3}{\pi} \right)^{1/3} [n(\mathbf{r})]^{1/3}. \quad (6.101)$$

V_x represents the Dirac-Slater exchange potential, differing from (6.33) because of the $3/2$ factor.

6.5.5 Correlation energy

A correction to Thomas-Fermi-Dirac approximation is to be considered to take into account the correlations between electrons with opposite spin. In other words, the density depletion occurring around a given particle is influenced by Coulomb repulsion: this correlates the electrons so that the position of each of them is not independent of the position of all the others. Indicating by ϵ_{xc} the exchange and correlation energy of the homogeneous gas, the density-potential relation (6.98) can be generalized into the form:

$$\mu = \frac{5}{3}C_k [n(\mathbf{r})]^{2/3} + V_n(\mathbf{r}) + V_e(\mathbf{r}) + \frac{\delta\epsilon_{xc}}{\delta n}. \quad (6.102)$$

The functional variation $(\delta\epsilon_{xc})/(\delta n)$ can be calculated within the local-density approximation, as discussed in Sec. 6.6.

6.5.6 Density-gradient corrections

The most important source of errors in the use of Thomas-Fermi-Dirac model is connected to the local-density approximation. This approximation is nicely applied

to weakly inhomogeneous electron gases, whereas for atoms and molecules both $V(\mathbf{r})$ and $n(\mathbf{r})$ are not slowly-varying functions of space. Since density gradients are related to the existence of currents, the question arises whether a local current-density approximation is to be perhaps formulated here, that might produce better approximations. This is indeed the case, as discussed in Sec. 6.11.

To conclude this section, one sees that all the non-classical contributions to the two-particle interaction can be represented as a one-particle potential. Though obtained within approximations, this result is more general. Two corrections need to be taken into account when using the Thomas-Fermi-Dirac approximation, that are related to exchange and correlation between opposite-spin particles and to the existence of current processes. All these puzzle pieces are fit together in the Density Functional Theory discussed in the rest of the Chapter, along with its extension to time-dependent processes.

6.6 The Density Functional Theory

The Density Function Theory (DFT) and its extensions to time-dependent systems and to special types of ground states, such as spin or superconducting systems, represents a key powerful advancement in conceptual, procedural, and factual knowledge. The DFT concept is based on the Hohenberg and Kohn theorem [12], stating that the ground state energy of a correlated system of many particles in an external potential is a functional $F[n(\mathbf{r})]$ of the density profile $n(\mathbf{r})$, so that all the static properties of a non-degenerate ground state of a quantum system can be described in terms of its density. The functional accounts in exact manner for the kinetic and exchange and correlation contribution to the ground-state energy. Based on this crucial result, Kohn and Sham [13] have demonstrated that the many-particle system can be mapped onto a suited and fictitious non-interacting system of particles moving in an effective potential $V_{\text{eff}}(\mathbf{r})$ depending on the whole density distribution $n(\mathbf{r}')$ everywhere, not just at position \mathbf{r} , and including the bare potential energy and the exchange and correlation contribution $V_{xc}(\mathbf{r})$.

Practical schemes are available to apply these concepts and calculate the predicted behavior of microscopic quantities that can be observed in atomic, molecular, solid-state physics. The DFT predictions are most often found to provide excellent agreement with experimental findings, sometimes in fact even outside the range of validity of the adopted approximation that can be inferred a priori. One among the most used and best performing approximations to the exchange and correlation potential is the Local Density Approximation (LDA), that in principle is valid under weakly inhomogeneity conditions, though in practice it often works beyond this limit. Here, $V_{xc}(\mathbf{r})$ is calculated by substituting to the density of the homogeneous gas the local density $n(\mathbf{r})$ of the inhomogeneous system, wherever it appears in the equation. This leads to the calculation of V_{xc} for the homogeneous system, that can be more easily performed by means of theoretical perturbative methods or by Quantum Monte Carlo simulational methods. In essence then, the static properties of

weakly inhomogeneous and strongly correlated quantum system can be determined in terms of a corresponding homogeneous non-interacting fictitious system.

The DFT can be applied to systems with a large number of particles, providing reliable results within 1% accuracies, with no special efforts. Modern calculations aimed to reach better accuracies are based on completely different techniques, such as numerical exact diagonalization of systems with a few particles.

The DFT has been applied to investigate the ground state structural and electronic properties of atoms, molecules, solids, and systems in reduced dimensions. A non exhaustive list of examples includes the determination of: geometrical properties such as bond lengths within 1-2% accuracy; elastic constants within 5-10%; dielectric and piezoelectric properties in response to external perturbations; lattice vibrations within a few percent accuracy; phase transitions, such as pressure-induced ones and the corresponding stability of phase diagrams [14].

Extensions of the DFT have been developed to treat atoms and molecules with nonzero spin, relativistic effects, paramagnetism and ferromagnetism in metals, finite-temperature systems, excited states, multi-component systems, superfluid systems such as liquid He^4 and trapped ultracold quantum atomic gases. The description of dynamical behavior requires a further extension to the Time-Dependent DFT (TD-DFT), that has been formulated by Vignale and Kohn [16, 17] in terms of the current density. Use of conservation laws and general symmetry arguments has been shown to lead to a quite interesting result: the equations of TD-DFT are in fact formally equivalent to phenomenological Navier-Stokes equations normally used to investigate fluid motion, where however all the physical quantities can be expressed in terms of their microscopic definitions and of current response functions [18]. These results have been extended to the investigation of superfluid dynamical systems [19, 20], where the Navier-Stokes equations result to be substituted by the Landau-Khalatnikov phenomenological equations that describe the dynamics of superfluids.

The rest of the Chapter is aimed to provide the basics of DFT for electron systems and a selection of applications to the determination of real systems. One section is finally devoted to briefly picture the conceptual framework that fits together the puzzle pieces collected along the textbook development, in particular: DFT and TD-DFT with crystal, electron and phonon structures, transport and optical susceptibilities, correlations in many-particle quantum systems, density and current-density response function theory.

6.6.1 Variational formulation of the theory

Consider a system of N particles at zero temperature, moving under the action of an external single-particle potential $V_e(\mathbf{r})$ and of their mutual internal interaction potential $U(\mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_{i < j} V(\mathbf{r}_i, \mathbf{r}_j)$. The system Hamiltonian is $H = T + U + V$, where $V(\mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_{i=1}^N V_e(\mathbf{r}_i)$. The corresponding non-degenerate ground-state normalized wavefunction $\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ is a unique functional of $V_e(\mathbf{r})$. This means

that the ground state is determined in univocal manner once $V_e(\mathbf{r})$ is given, with the density given by (6.1).

The fundamental theorem of DFT states that between $V_e(\mathbf{r})$ and $n(\mathbf{r})$ a one-to-one correspondence exists, to within an additive constant for $V_e(\mathbf{r})$. The proof is by reduction to absurd. Assume that a second and different potential $\tilde{V}_e(\mathbf{r})$ exists, to which a different ground state $\tilde{\Psi}$ be associated leading though to the same $n(\mathbf{r})$. Of course $\tilde{\Psi} \neq \Psi$ because they are eigenstates of different Hamiltonians and thus satisfy different Schrödinger equations, unless $V_e(\mathbf{r}) - \tilde{V}_e(\mathbf{r}) = \text{constant}$. Consider the wavefunctions labeled by Ψ and $\tilde{\Psi}$, the Hamiltonians by H and \tilde{H} , and the corresponding ground-state energies by E and \tilde{E} , respectively. Since E is the energy of the ground state of H while $\tilde{\Psi}$ is not its eigenstate, the Rayleigh-Ritz variational principle implies that

$$E = \langle \Psi | H | \Psi \rangle < \langle \tilde{\Psi} | H | \tilde{\Psi} \rangle = \langle \tilde{\Psi} | \tilde{H} + V - \tilde{V} | \tilde{\Psi} \rangle. \quad (6.103)$$

Since the two states originate from the same density $n(\mathbf{r})$, one obtains

$$E < \tilde{E} + \int [V_e(\mathbf{r}) - \tilde{V}_e(\mathbf{r})] n(\mathbf{r}) d\mathbf{r}. \quad (6.104)$$

In similar manner, swapping the quantities with and without superscript one obtains

$$\tilde{E} < E + \int [\tilde{V}_e(\mathbf{r}) - V_e(\mathbf{r})] n(\mathbf{r}) d\mathbf{r}. \quad (6.105)$$

Summing the two relations above on gets the absurd result

$$E + \tilde{E} < E + \tilde{E}. \quad (6.106)$$

In conclusion, the density function associated to a non-degenerate ground state determined by an external potential $V_e(\mathbf{r})$ cannot be reproduced by the ground-state density determined by a different $\tilde{V}_e(\mathbf{r})$, unless $V_e(\mathbf{r})$ and $\tilde{V}_e(\mathbf{r})$ differ by a constant. The reverse is also true. Indeed, since $n(\mathbf{r})$ determines the potential $V_e(\mathbf{r})$ and the latter fixes the Hamiltonian H , the density $n(\mathbf{r})$ determines as well the ground state of H , provided it is non-degenerate. Therefore:

Concept

On quite general grounds, the Hohenberg and Kohn theorem states that a one-to-one correspondence exists between the density $n(\mathbf{r})$ of an interacting system and the external potential $V_e(\mathbf{r})$ acting on it.

A few remarks are in order. The non-degeneracy requirement for the ground state is crucial, though it can be appropriately removed. The theorem holds as well for non-interacting systems. Mermin has extended the theorem demonstration to finite-temperature systems, using the free energy within the grand-canonical ensemble: this leads to the Hohenberg-Kohn-Mermin theorem [15].

Since Ψ is a unique functional of $n(\mathbf{r})$, the density functional is also well defined

$$F[n(\mathbf{r})] = \langle \Psi | T + U | \Psi \rangle, \quad (6.107)$$

that represents the sum of kinetic and interaction energies. This is a universal functional, valid whatever the external potential or the number N of particles might be, as it is obvious since N itself is a functional $n(\mathbf{r})$. The determination of this universal functional is the key issue of DFT.

6.6.2 Kohn-Sham equations

The density functional corresponding to the ground state of a system in the presence of an external potential is

$$E_g[n(\mathbf{r})] = T[n(\mathbf{r})] + \int V_e(\mathbf{r})n(\mathbf{r})d\mathbf{r} + \frac{1}{2} \int d\mathbf{r}d\mathbf{r}' V(\mathbf{r}, \mathbf{r}')n(\mathbf{r})n(\mathbf{r}') + E_{xc}[n(\mathbf{r})]. \quad (6.108)$$

On the right-hand side of (6.108):

Properties

P1. $T[n(\mathbf{r})]$ represents the functional of kinetic energy that corresponds to the ideal non-interacting system .

P2. $\int V_e(\mathbf{r})n(\mathbf{r})d\mathbf{r}$ is the contribution of the external field .

P3. $(1/2) \int d\mathbf{r}d\mathbf{r}' V(\mathbf{r}, \mathbf{r}')n(\mathbf{r})n(\mathbf{r}')$ represents the contribution of the potential energy between each pair of particles

P4. $E_{xc}[n(\mathbf{r})]$ represents the exchange and correlation contribution, defined as all what is to be added to the above P1-P3 terms to obtain the exact ground state energy $E_g[n(\mathbf{r})]$.

As already discussed in Sec. 6.4.2, $E_{xc}[n(\mathbf{r})]$ accounts for the excess energy to be added to both the kinetic $T[n(\mathbf{r})]$ and the pair potential pieces to get the real kinetic and interaction energies.

By definition, $E_g[n(\mathbf{r})]$ is to be stationary with respect to variations of $n(\mathbf{r})$, with the constraint that the total number of particles be conserved. Therefore, one has

$$\frac{\delta E_g[n(\mathbf{r})]}{\delta n(\mathbf{r})} = \mu, \quad (6.109)$$

whose solution determines the ground state density profile $n_0(\mathbf{r})$ and chemical potential μ . On the other hand, the expression equivalent to (6.109) for a non-interacting system in the external potential $V_e(\mathbf{r})$ reads

$$\frac{\delta T[n(\mathbf{r})]}{\delta n(\mathbf{r})} + V_e(\mathbf{r}) = \mu. \quad (6.110)$$

Thus, according to the Kohn and Sham scheme

Concept

The interacting system can be mapped onto an equivalent and effective non-interacting one, characterized by the effective single-particle potential

$$V_{\text{eff}}(\mathbf{r}) = V_e(\mathbf{r}) + \int d\mathbf{r}' V(\mathbf{r}, \mathbf{r}') n(\mathbf{r}') + \frac{\delta E_{xc}[n(\mathbf{r})]}{\delta n(\mathbf{r})}. \quad (6.111)$$

The ground-state energy E_g and corresponding density $n_0(\mathbf{r})$ can therefore be calculated after solving the set of Schrödinger equations

$$-\frac{\hbar^2}{2m} \nabla^2 \psi_i(\mathbf{r}) + V_{\text{eff}}(\mathbf{r}) \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r}) \quad (6.112)$$

for the fictitious single-particle orbitals $\psi_i(\mathbf{r})$. In particular, the ground-state density profile is immediately given by $n_0(\mathbf{r}) = \sum_{i=1}^N |\psi_i(\mathbf{r})|^2$ with the sum extended to all the occupied fictitious states. The parameters ε_i satisfy the relation

$$\begin{aligned} \sum_i \varepsilon_i &= \sum_i \langle \psi_i(\mathbf{r}) \left[-\frac{\hbar^2}{2m} \nabla^2 + V_{\text{eff}}(\mathbf{r}) \right] \psi_i(\mathbf{r}) \rangle \\ &= T[n_0(\mathbf{r})] + \int d\mathbf{r} V_{\text{eff}}(\mathbf{r}) n_0(\mathbf{r}). \end{aligned} \quad (6.113)$$

By means of (6.108), the ground-state energy can thus be cast in the form

$$\begin{aligned} E_g &= \sum_i \varepsilon_i - \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' V(\mathbf{r}, \mathbf{r}') n_0(\mathbf{r}) n_0(\mathbf{r}') \\ &\quad + E_{xc}[n_0(\mathbf{r})] - \int d\mathbf{r} V_{xc}(\mathbf{r}) n_0(\mathbf{r}). \end{aligned} \quad (6.114)$$

Here, the exchange and correlation term is given by

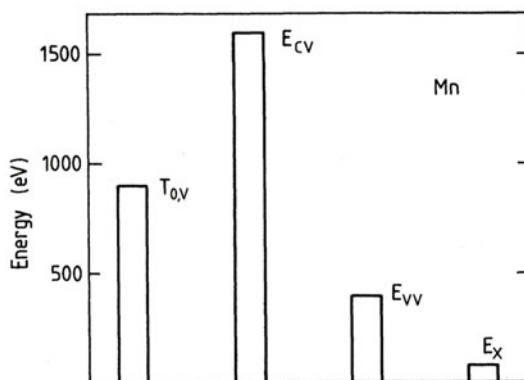
$$V_{xc}(\mathbf{r}) \equiv \frac{\delta E_{xc}[n(\mathbf{r})]}{\delta n(\mathbf{r})} \Big|_{n(\mathbf{r})=n_0(\mathbf{r})}. \quad (6.115)$$

Relations (6.112) represent the self-consistent Kohn and Sham equations for the fictitious single-particle orbitals and energies. Due to their link with the exact physical density $n_0(\mathbf{r})$, the single-particle Kohn and Sham orbitals $\psi_i(\mathbf{r})$ can be considered as optimized in density, whereas in the Hartree-Fock case they are optimized in total energy.

Notice that independent knowledge of $n(\mathbf{r})$ as it can be determined from experiments, or from numerical accurate or else exact calculations in the case of very small systems, $V_{\text{eff}}(\mathbf{r})$ and therefore $V_{xc}(\mathbf{r})$ can be obtained. This possibility is discussed in Sec. 6.7.3 for He atom and Li^+ and Be^{++} ions.

Omitting the last term in (6.108), the variational minimization procedure of E_g with respect to $n(\mathbf{r})$ would lead to the Hartree solution. The great advantage of separating the energy functional as in (6.108) is at a glimpse apparent in Fig. 6.6, where the contribution of valence electrons to the total energy of a manganese atom is displayed: the exchange E_x and correlation energy are quite small, the latter being not even visible in the graph. Thus, the kinetic term $T_{0,V}$ of the valence electrons, the electrostatic interaction term E_{CV} between valence and core electrons, and the valence-electron electrostatic interaction term E_{VV} can be calculated exactly, reserving the effort of approximate calculations only to the much smaller exchange and correlation pieces.

Fig. 6.6 Density Functional Theory. Advantage of separating the exchange and correlation terms from the total energy. Calculated contribution of valence electrons to the total energy of a manganese atom. Exchange E_x , kinetic energy $T_{0,V}$ of valence electrons, valence-core E_{CV} and valence-valence E_{VV} electrostatic interaction energy are indicated. The correlation energy is even smaller than E_x and not visible. Reprinted with permission from [21]. Copyright (1989) by the American Physical Society



6.6.3 Meaning of the Kohn-Sham eigenvalues and eigenstates

The solution of Kohn-Sham equations leads to the determination of occupied energy levels with ϵ_i and $i = 1, \dots, N$, as well as non-occupied ones with ϵ_i and $i = N + 1, \dots$. The question arises whether the values of the occupied levels represent the energy required to excite one particle to the continuum. The answer is negative, since as already noticed the Kohn and Sham eigenvalues and eigenstates are fictitious, not bearing any physical meaning. One exception exists to the latter

statement, that is the case of an infinite system whose higher Kohn-Sham occupied energy states are extended. In this case, the eigenvalue ϵ_N of the highest occupied state coincides with the value of the chemical potential. This result represents the analogue of Koopman theorem in the case of the density functional.

A second relevant question arises in regard to solids, that are periodical infinitely extended systems. In this case, the Kohn and Sham eigenfunctions carry a crystal momentum \mathbf{k} . A Fermi surface can also be defined by means of $\epsilon_{\mathbf{k}} = \mu$ at $T = 0$, with μ the chemical potential of the real system: μ would be characterized by the property that the total number of states with energy not exceeding μ , coincides with the total number of electrons N . The question is whether the Fermi surface calculated from the Kohn-Sham equations is related and how to the Fermi surface of the real system. This is not necessarily so, or at least no general proof exists. It can only be stated by construction that the two Fermi surfaces bound the same volume in \mathbf{k} space, so that in the special case of a uniform system having by definition a spherical Fermi surface, the real and fictitious Fermi surfaces do necessarily coincide.

6.7 Approximations for the exchange-correlation functional

The problem remains on how the exchange and correlation energy functional $E_{xc}[n(\mathbf{r})]$ can be calculated in practice. This is discussed in the present section.

6.7.1 Local-Density Approximation (LDA)

A most often effective and quite popular approximation is that of local density (LDA). Consider a system whose density is a slowly-varying function of space. Under this condition, one can reasonably assume that the density $n(\mathbf{r})$ be slightly differing at each location \mathbf{r} from the density of the homogeneous system. In this case one can write [13]

$$E_{xc}[n(\mathbf{r})] = \int d\mathbf{r} \epsilon_{xc}(n(\mathbf{r})) n(\mathbf{r}),$$

where $\epsilon_{xc}(n(\mathbf{r}))$ is the exchange and correlation energy per particle of the uniform gas with the density value corresponding to $n(\mathbf{r})$. In practice, once the function $\epsilon_{xc}(n)$ is calculated by means of theoretical or simulational methods for the homogeneous system, the substitution $n \rightarrow n(\mathbf{r})$ is performed. Notice that $\epsilon_{xc}(n(\mathbf{r}))$ here is a function of $n(\mathbf{r})$ and not a functional, as in $\epsilon_{xc}[n(\mathbf{r})]$. Therefore, $V_{xc}(\mathbf{r})$ depends on the value of the density at position \mathbf{r} alone. Indeed,

$$V_{xc}(\mathbf{r}) = \frac{d}{dn} \{ \epsilon_{xc}(n(\mathbf{r})) n(\mathbf{r}) \} \equiv \mu_{xc}(n(\mathbf{r})) \quad (6.116)$$

where $\mu_{xc}(n(\mathbf{r}))$ is the exchange and correlation contribution to the chemical potential of a uniform system. In LDA the expression (6.114) becomes:

$$E_g \simeq \sum_i \varepsilon_i - \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' V(\mathbf{r}, \mathbf{r}') n(\mathbf{r}) n(\mathbf{r}') + \int d\mathbf{r} [\varepsilon_{xc}(n(\mathbf{r})) - \mu_{xc}(n(\mathbf{r}))] n(\mathbf{r}). \quad (6.117)$$

In the case of a many-electron system for example, the exchange contribution to $\varepsilon_{xc}(n(\mathbf{r}))$ could be approximated to be the Hartree-Fock result (6.94):

$$\begin{aligned} \varepsilon_{xc}(n(\mathbf{r})) &= - \left[\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} e^2 \right] n(\mathbf{r})^{1/3} = \\ &= - \frac{3}{4} \left(\frac{9}{4\pi^2} \right)^{1/3} \frac{e^2}{r_s} = - \frac{0.458}{r_s} e^2 \text{ Ry}. \end{aligned} \quad (6.118)$$

For better results, perturbative theoretical methods are often substituted by numerical methods such as exact diagonalization of systems with few particles or Quantum Monte Carlo simulations. Pioneering simulational results for the electron gas are due to Ceperley and Ceperley and Alder [22, 23], that are at the 1-2% accuracy level. Fits of the ground-state energy as a function of density for the homogeneous system are usually provided once and for all, to be used as essential ingredients for DFT calculations.

LDA is by definition exact when operated on a homogeneous system. The approximation is expected to provide accurate results as long as the real system is characterized by weak inhomogeneity. Surprisingly enough, LDA is a posteriori and in practice seen to provide accurate results also beyond this limits of validity in principle. For example, calculations of the dissociation energy of molecules, ionization energies of atoms, and cohesive energies in solids is within a 10% accuracy, while bond distances are reproduced to the best 1% accuracy. In their seminal paper, Kohn and Sham [13] stated that LDA could not be adequate to describe chemical bonds in molecules, where the condition of slow spatial variation is broken precisely in those regions of major overlap of charge densities from different atoms. The reason why LDA yields accurate results in spite of the degree of inhomogeneity, is that it satisfies the sum rule (6.25) and therefore embodies an essential system peculiarity.

The accuracies of LDA estimates for exchange energies are of the order of 10%, whereas those for correlation energies typically lead to values a factor 2 larger. Since correlation energies are typically smaller than exchange ones, the two errors often cancel out to some extent.

LDA is completely inadequate for all those systems for which the approximation of homogeneous system is not correct. A typical example is the Wigner crystal ground state. A second example is given by the Van der Waals interaction introduced in Chap. 1, for which the overlap of molecule wavefunctions composing the crystal is very small. In fact, DFT can still be used to treat these kind of systems, though different and suited approximations to E_{xc} are to be adopted.

6.7.2 Connecting the exchange-correlation functional and the exchange and correlation hole

Considered the discussions around the exchange and correlation hole and around the exchange and correlation energy functional, the question arises on how the two concepts are connected.

Kohn and Sham equations are based on the crucial mapping between the interacting and the non-interacting system: consequently, the exchange and correlation energy is defined as that error that is made by performing this mapping and by neglecting the quantum contributions to the pairs interactions. The concept of exchange and correlation hole as resulting from (6.24) is of great help here. The following considerations are specialized to Coulomb systems, though they may be generalized.

The exchange and correlation hole satisfies the normalization condition (6.25), physically related to the total screening that an electron located at position \mathbf{r} experiences. The hole is also localized, due to the combined effect of Pauli exclusion principle and of electron-electron interactions. Of course, in accord to Hohenberg and Kohn theorem, it is a functional of density.

The direct connection between $E_{xc}[n(\mathbf{r})]$ and the exchange and correlation hole can be appropriately modified. This is:

$$E_{xc}[n(\mathbf{r})] = \frac{1}{2}e^2 \int d\mathbf{r} n(\mathbf{r}) \int d\mathbf{r}' \frac{\bar{n}_{xc}(\mathbf{r}, \mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|}, \quad (6.119)$$

where $\bar{n}_{xc}(\mathbf{r}_1, \mathbf{r}_2)$ is defined by means of

$$\bar{n}_{xc}(\mathbf{r}_1, \mathbf{r}_2) = \int_0^1 d\lambda n_{xc}(\mathbf{r}_1, \mathbf{r}_2, \lambda) \quad (6.120)$$

and the parameter λ , with $0 \leq \lambda \leq 1$, scales as the strength of the interaction, so that the external potential becomes a suited function of λ and the ground-state density corresponding to the λ -dependent Hamiltonian H_λ , be equal to the real physical density.

The following properties are conveniently remarked for Coulomb systems:

Properties

P1. Since exchange and correlation terms tend to vanish while $|\mathbf{r} - \mathbf{r}'| \rightarrow +\infty$ and thus $\bar{n}_{xc}(\mathbf{r}, \mathbf{r}' - \mathbf{r}) \rightarrow 0$, the difference between electrostatic and exchange and correlation energies can be viewed as an approximate separation between the long- and short-range part of the interaction. One can therefore argue that the total interaction energy be less sensitive against density variations, since the long-range part can be exactly calculated.

P2. If the bare interaction is isotropic, as in the Coulomb case, angular averages can be conveniently performed, leading to

$$E_{xc} [n(\mathbf{r})] = -\frac{1}{2}e^2 \int d\mathbf{r} n(\mathbf{r}) \left\langle \frac{1}{R} \right\rangle_{\mathbf{r}}, \quad (6.121)$$

with

$$\frac{1}{R_{xc}(\mathbf{r})} \equiv \left\langle \frac{1}{R} \right\rangle_{\mathbf{r}} = -e^2 \int d\mathbf{R} \frac{\bar{n}_{xc}(\mathbf{r}, \mathbf{R})}{R}. \quad (6.122)$$

This result shows that the exchange and correlation energy is weakly dependent on the details of $n_{xc}(\mathbf{r}, \mathbf{r}' - \mathbf{r})$, provided that the sum rule (6.25) is satisfied.

Fig. 6.7 display the exact profile of the exchange hole for a neon atom, as compared with its LDA approximation for an electron at two different distances r along the electron-nucleus direction. This example is significant, since in atoms exchange can be calculated exactly. Besides, it normally prevails on correlation effects. Therefore, the latter can be neglected and the exchange hole has the same characteristics of the exchange and correlation hole, in particular it satisfies the sum rule (6.25). In particular, in the left panels of the two figures it is seen that the exact hole is always very localized and centered on the nucleus. The LDA hole instead, that is here calculated from the Hartree-Fock expression (6.94), is more extended, spherically symmetric, and centered on the electron. Only when a spherical angular average is performed, the agreement of LDA with the exact calculation improves, as it is visible in the right panels. Since the plotted curves are related to $R\bar{n}_{xc}(\mathbf{r}, \mathbf{R})$, the area below each curve is directly proportional to the corresponding exchange energy density as in (6.121). It is also evident how the cancellation between the errors made in the estimates of both exchange and correlation contributions is most effective at intermediate and large distances from the hole center. The cancellation originates from the sum rule (6.23), stating that the area below the curve $-4\pi R^2 \bar{n}_{xc}(\mathbf{r}, \mathbf{R})$ is to sum up to -1 in both cases. Therefore, errors in calculating the details of this curve within LDA as compared to the exact calculation, are to compensate to each other so that the two curves, multiplied by R and integrated over, give the same number.

All the above considerations are not substantially modified when correlation effects are included. The latter essentially reduce the spatial extension of the pair correlation function. As a consequence, the LDA exchange and correlation hole gets a more realistic localization around the particle. This argument suggests to adopt strategies where exchange and correlation terms are simultaneously treated.

In conclusion, the large success of LDA can be summarized to be due to the following reasons. First, a systematic cancellation occurs between the errors resulting from the calculation of exchange and correlation contributions, as a consequence of the sum rule expressing particle-density conservation. Second, the LDA exchange and correlation energy can be improved by spherical angle averages, whenever useful.

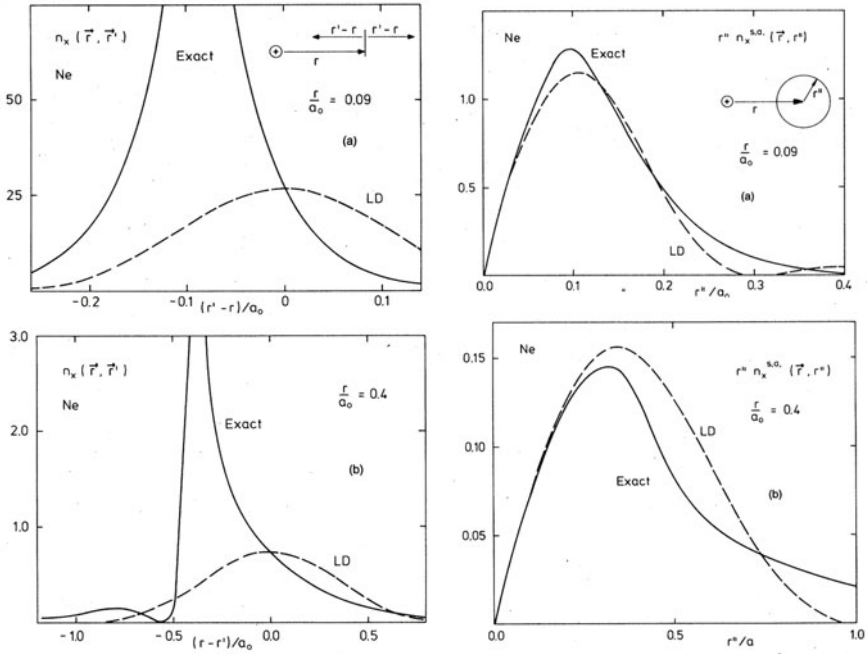


Fig. 6.7 Exchange and correlation hole profile in atoms and comparison with LDA results. The hole spatial profile of a Ne atom around an electron at two different positions along the electron-nucleus direction, as labeled in the panels. Top panels: $r/a_B = 0.09$. Bottom panels: $r/a_B = 0.4$. Solid lines: exact calculation. Dashed lines: LDA calculation. The non angle-averaged and the spherically averaged results are shown in the left and right panels, respectively. Reprinted with permission from [21]. Copyright (1989) by the American Physical Society

6.7.3 Selected exact results

As anticipated, the knowledge of the exact wavefunction allows the calculation of the density and of the exchange and correlation $E_{xc}[n(\mathbf{r})]$ and potential $V_{xc}(\mathbf{r})$. This turns out to be useful in understanding the degree of reliability of the adopted approximations for the exchange and correlation functional.

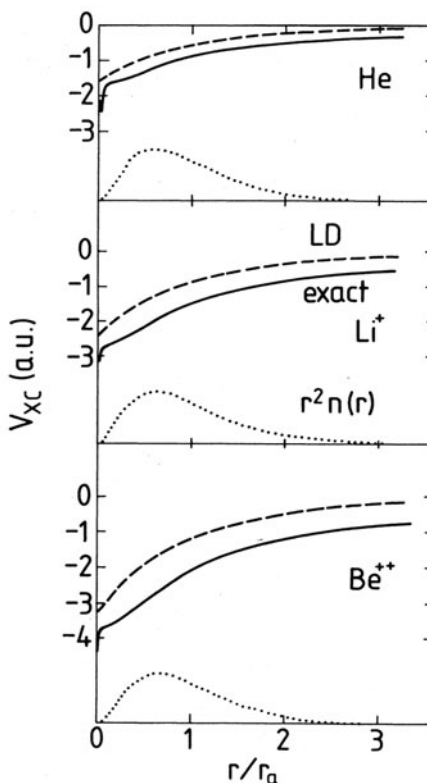
The example of He is now given. He possesses one single doubly occupied orbital with eigenvalue ε_1 and wavefunction ψ_1 , so that $\varphi(\mathbf{r}) = |\psi_1|^2 = n(\mathbf{r})/2$. Inversion of (6.112) leads to the following expression for $V_{xc}(\mathbf{r})$:

$$V_{xc}(\mathbf{r}) = \frac{\nabla^2 \varphi(\mathbf{r})}{2\varphi(\mathbf{r})} - V_e(\mathbf{r}) - U_{es} - \varepsilon_1, \quad (6.123)$$

where U_{es} is given by (6.16). In atoms with more than two electrons, $V_{xc}(\mathbf{r})$ can be parametrized and the parameters be fitted from the exact density profile. Though $E_{xc}[n(\mathbf{r})]$ is accurately enough described within LDA with accuracies below 10 %, the accuracy on $V_{xc}(\mathbf{r})$ is expected to be worst. This is visible in Fig. 6.8, where the exact and LDA calculations of $V_{xc}(\mathbf{r})$ are compared for the cases of He, Li^+

and Be^{++} . Though the error on $V_{xc}(\mathbf{r})$ implies a large error on the eigenvalues, one sees that the deviations of the LDA exchange and correlation potential from the exact result is to good approximation independent of the distance from the nucleus, thereby reducing the errors on the density profiles.

Fig. 6.8 Exchange and correlation potential for selected atoms and ions and comparison with LDA results. V_{xc} is displayed in atomic units as a function of distance scaled to the atomic radius r_a . The cases of He, Li^+ and Be^{++} are shown as in the legend, with $r_a = 0.929, 0.573$ and 0.414 a.u., respectively. Solid lines: exact calculation. Dashed lines: LDA calculation. Dotted line: the corresponding density profile. Reprinted with permission from [21]. Copyright (1989) by the American Physical Society



6.8 Connection between DFT and response function theory

In view of the concepts developed in Sec. 6.6, it is not surprising to state that DFT is strictly connected with the theory of linear density response and of the dielectric function developed in Secs. 5.3 and 5.2.1 of Chap. 5. At the end of this Chapter, it is also discussed that the time-dependent version of DFT is strictly connected with current linear response, whose concept has been exploited in Chap. 4 to calculate the transport coefficients. This section is aimed to establish the formal connection between DFT and density response to linear order, though the argument could be extended to non-linear response function theory as well.

Consider the effective DFT potential (6.111)

$$V_{\text{eff}}(\mathbf{r}) = V_e(\mathbf{r}) + \int d\mathbf{r}' V(\mathbf{r}, \mathbf{r}') n(\mathbf{r}') + \frac{\delta E_{xc}[n(\mathbf{r})]}{\delta n(\mathbf{r})} \Big|_{n(\mathbf{r})=n_0(\mathbf{r})}, \quad (6.124)$$

In the presence of a small variation of the external potential $\delta V_e(\mathbf{r})$, a small induced density fluctuation $\delta n(\mathbf{r})$ takes place. Expression (6.124) can be expanded to linear order in δn , so that one has

$$\delta V_{\text{eff}}(\mathbf{r}) \simeq \delta V_e(\mathbf{r}) + \int d\mathbf{r}' \delta n(\mathbf{r}') V(\mathbf{r}, \mathbf{r}') + \int d\mathbf{r}' \delta n(\mathbf{r}') \frac{\delta^2 E_{xc}[n(\mathbf{r})]}{\delta n(\mathbf{r}) \delta n(\mathbf{r}')} \Big|_{n(\mathbf{r})=n_0(\mathbf{r})}. \quad (6.125)$$

One can define the kernel

$$f_{xc}(\mathbf{r}, \mathbf{r}') = \frac{\delta^2 E_{xc}[n(\mathbf{r})]}{\delta n(\mathbf{r}) \delta n(\mathbf{r}')} \Big|_{n(\mathbf{r})=n_0(\mathbf{r})}, \quad (6.126)$$

that determines the exchange and correlation potential $\delta E_{xc}[n(\mathbf{r})]/\delta n(\mathbf{r})$. The kernel f_{xc} can be viewed as the factor that shapes the exchange and correlation hole.

The static density fluctuation can also be calculated as the response of the fictitious Kohn-Sham system to the effective potential δV_{eff} , that is

$$\delta n(\mathbf{r}) = \int d\mathbf{r}' \chi_{KS}(\mathbf{r}, \mathbf{r}') \delta V_{\text{eff}}(\mathbf{r}'). \quad (6.127)$$

with χ_{KS} the Kohn-Sham response function corresponding to the fictitious non-interacting system. χ_{KS} is therefore built from the fictitious single-particle orbitals, that is:

$$\chi_{KS}(\mathbf{r}, \mathbf{r}') = \sum_{i,j} (f_i - f_j) \frac{\psi_i^*(\mathbf{r}) \psi_j(\mathbf{r}') \psi_j^*(\mathbf{r}) \psi_i(\mathbf{r}')}{\epsilon_i - \epsilon_j}, \quad (6.128)$$

with $f_{i,j}$ the occupation numbers of the particles in the quantum energy states labeled by i, j .

By inspection, expressions (6.125) and (6.127) give

$$f_{xc}(\mathbf{r}, \mathbf{r}') = \chi_{KS}^{-1}(\mathbf{r}, \mathbf{r}') - \chi^{-1}(\mathbf{r}, \mathbf{r}') - V(\mathbf{r}, \mathbf{r}'). \quad (6.129)$$

In terms of the dielectric function instead, one first notices that

$$\delta V_{\text{eff}}(\mathbf{r}) = \int d\mathbf{r}' \frac{1}{\epsilon(\mathbf{r}, \mathbf{r}')} \delta V_e(\mathbf{r}'), \quad (6.130)$$

connecting the total effective potential $V_{\text{eff}}(\mathbf{r}, t)$ to the external field $V_e(\mathbf{r})$. Expression (6.130) can be cast in the form

$$\delta V_{\text{eff}}(\mathbf{r}) = \int d\mathbf{r}' \left[\frac{1}{\epsilon(\mathbf{r}, \mathbf{r}')} - \delta(\mathbf{r} - \mathbf{r}') \right] \delta V_e(\mathbf{r}') + \delta V_e(\mathbf{r}).$$

By inspection, the comparison with (6.125) yields

$$\int d\mathbf{r}' \left[\frac{1}{\epsilon(\mathbf{r}, \mathbf{r}') - \delta(\mathbf{r} - \mathbf{r}')} \right] \delta V_e(\mathbf{r}') = \int d\mathbf{r}' [V(\mathbf{r}, \mathbf{r}') + f_{xc}(\mathbf{r}, \mathbf{r}')] \delta n(\mathbf{r}'). \quad (6.131)$$

Using the expression of $\delta n(\mathbf{r}')$ in terms of the external potential leads to first order to

$$\int d\mathbf{r}' \left[\frac{1}{\epsilon(\mathbf{r}, \mathbf{r}') - \delta(\mathbf{r} - \mathbf{r}')} \right] \delta V_e(\mathbf{r}') = \int d\mathbf{r}' d\mathbf{r}'' [V(\mathbf{r}, \mathbf{r}') + f_{xc}(\mathbf{r}, \mathbf{r}')] \chi_{KS}(\mathbf{r}', \mathbf{r}'') \delta V_e(\mathbf{r}'')$$

By performing the exchange $\mathbf{r}' \leftrightarrow \mathbf{r}''$ one finds

$$\frac{1}{\epsilon(\mathbf{r}, \mathbf{r}') - \delta(\mathbf{r} - \mathbf{r}')} = \int d\mathbf{r}'' [V(\mathbf{r}, \mathbf{r}'') + f_{xc}(\mathbf{r}, \mathbf{r}'')] \chi_{KS}(\mathbf{r}'', \mathbf{r}'), \quad (6.132)$$

In the homogeneous system, space Fourier transform can be operated, leading to

$$\frac{1}{\epsilon(q) - 1} = [v(q) + f_{xc}(q)] \chi_{KS}(q), \quad (6.133)$$

where $v(q) = 4\pi e^2/q^2$ for a charged Coulomb system. In terms of the response functions, one has

$$\chi(q) = \frac{\chi_{KS}(q)}{1 - v(q)\chi_{KS}(q) - f_{xc}(q)\chi_{KS}(q)}. \quad (6.134)$$

In essence,

Concept

DFT, response and dielectric function theories are strictly connected: effective and xc potentials calculated at the local density within the DFT, are related to the actual response function of the system and to a Kohn and Sham response function built up from the fictitious single-particle orbitals.

6.9 Comparison with other methods

After having set the formal development of the DFT and its connection with response and dielectric function theory, it is useful to revert back to the results already known from the approximate models discussed so far. In this manner, the relevance of DFT in extending the general concepts is better caught. The comparison with Hartree, Hartree-Fock, and Thomas-Fermi models is discussed in this section.

6.9.1 Comparison with Hartree and Hartree-Fock models

To compare DFT with the Hartree and Hartree-Fock methods, the separation is conveniently operated of $E_{xc}[n(\mathbf{r})]$ between the exchange and correlation contributions. The energy functional (6.108) becomes:

$$E_v[n(\mathbf{r})] = T_s[n(\mathbf{r})] + \int V_e(\mathbf{r})n(\mathbf{r})d\mathbf{r} + \frac{1}{2} \int d\mathbf{r}d\mathbf{r}' V(\mathbf{r}, \mathbf{r}')n(\mathbf{r})n(\mathbf{r}') + E_x[n(\mathbf{r})] + E_c[n(\mathbf{r})], \quad (6.135)$$

where $T_s[n(\mathbf{r})]$ is the kinetic energy of the non-interacting system with density $n(\mathbf{r})$, $E_x[n(\mathbf{r})]$ is the exchange energy as determined from a Hartree-Fock calculation, and $E_c[n(\mathbf{r})]$ is the correlation term.

Hartree and Hartree-Fock equations are obtained from (6.135) after eliminating $E_x[n(\mathbf{r})] + E_c[n(\mathbf{r})]$ or else $E_c[n(\mathbf{r})]$, respectively. The Lindhard approximation, also referred to as Random Phase Approximation, corresponds to set $f_{xc} = 0$.

The Hartree-Fock equations have been seen to embody the exchange term $E_x[n(\mathbf{r})]$ in exact manner. However, this implies the solution of a set of coupled integral-differential equations. The DFT instead, relaxes the accuracy on this term: according to practice, the error made is much smaller than that made within the Hartree-Fock scheme after neglecting the correlations. The advantage is that an ordinary set of differential equations is now at hand with DFT, that is of much easier solution than in the case of Hartree-Fock equations and only slightly more complicated than the Hartree ones. The Hartree-Fock equations can be simplified along this direction, by means of the LDA for $E_x[n(\mathbf{r})]$:

$$E_x[n(\mathbf{r})] = \int \varepsilon_x(n(\mathbf{r}))n(\mathbf{r})d\mathbf{r}. \quad (6.136)$$

In the case of interacting electrons, $\varepsilon_x(n(\mathbf{r}))$ is given by (6.118) and one obtains:

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' - \frac{1}{\pi} e^2 [3\pi^2 n(\mathbf{r})]^{1/3} \right\} \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r}), \quad (6.137)$$

whose exchange term is a factor 3/2 smaller than the original result from Slater and represents a more accurate approximation.

In any event,

Concept

The results obtained within the DFT for both the correlation and the exchange energies show up to be in general more accurate and realistic than those obtained within Hartree-Fock calculations.

6.9.2 Comparison with the Thomas-Fermi model

The original formulation of the Thomas-Fermi model can be derived as a special case of the DFT. To this aim, one has to eliminate the exchange and correlation terms in the functional and assume that the kinetic energy be that of a system of non-interacting electrons at the local density (6.61). This approximation holds only when the system density is slowly-varying and is exact for a homogeneous electron system. The DFT instead treats the kinetic energy in exact manner, with inclusion of the excess part.

As discussed in Sec. 6.5.2, the Thomas-Fermi method can be generalized to include exchange and correlation terms while keeping the local approximation for the kinetic energy. Consider the case of interacting electrons. Calculations of the kinetic energy within the Thomas-Fermi model imply that the charge density be infinite in the vicinity of the nucleus, so that the oscillations characterizing the atomic shell structure cannot be reproduced. These problems are resolved within the DFT. Indeed, first of all in the vicinity of nuclei the electronic density is high: exchange and correlation energies are therefore negligible in comparison to the kinetic term and thus, even though the validity conditions for the LDA are not satisfied, the DFT provides a satisfactory description. Second, in the central atomic region one expects that $n(\mathbf{r})$ be not rapidly varying, so that the local approximation for $\epsilon_{xc}[n(\mathbf{r})]$ becomes reasonable. Last, in the regions outside atoms and where the overlap molecular wavefunctions is large, the LDA is not very effective. Nonetheless, the chemical bond is nicely described since it is related to the exchange and correlation terms alone, and spherical averaging works well as already discussed.

6.10 Applications of the DFT

6.10.1 Calculation of band structures: the cases of Si, GaAs and Graphene

In Chap. 2 the semi-empirical pseudopotential method has been used to calculate the valence and conduction band structures of Si, GaAs, and Graphene crystals. There, it has been shown that the introduction of a few parameters is needed to make the results more realistic in comparison with experimental findings. Fig. 6.9 shows the results of DFT calculations. Comparing with those obtained in Chap. 2, one sees that the two are consistent, though DFT represents an ab initio calculation, not requiring any empirical parameter.

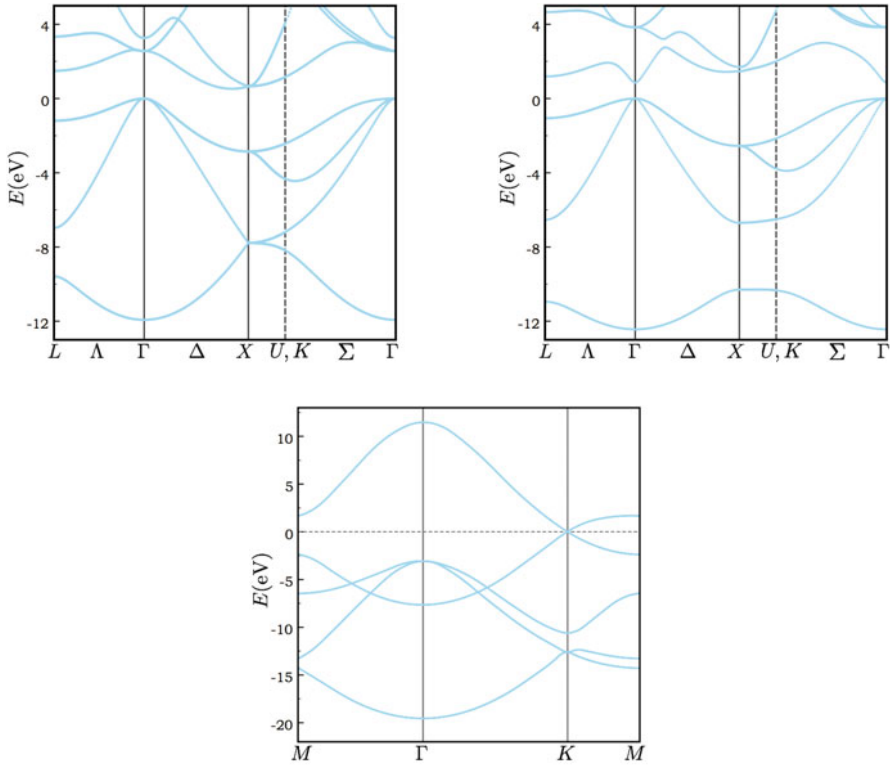


Fig. 6.9 Applications of DFT. Band structure of Si, GaAs and Graphene as calculated by means of DFT

6.10.2 Calculation of crystal dielectric functions

The total $\phi(\mathbf{k}, \omega)$ and external $\phi_{\text{ext}}(\mathbf{k}, \omega)$ potentials have been seen to be related to the dielectric function $\epsilon(\mathbf{k}, \omega)$, via $\epsilon(\mathbf{k}, \omega)\phi(\mathbf{k}, \omega) = \phi_{\text{ext}}(\mathbf{k}, \omega)$. Charge densities in crystals are not slowly-varying functions, thus (6.130) is better used, that after operating a Fourier transform becomes:

$$\phi(\mathbf{q} + \mathbf{G}', \omega) = \sum_{\mathbf{G}} \epsilon^{-1}(\mathbf{q} + \mathbf{G}', \mathbf{q} + \mathbf{G}, \omega) \phi_{\text{ext}}(\mathbf{q} + \mathbf{G}, \omega). \quad (6.138)$$

The terms $\mathbf{G} \neq \mathbf{G}'$ yield the local field effects. Their contribution is significant whenever the charge distribution is not slowly-varying, as it occurs in crystals. The calculation of $\epsilon(\mathbf{q} + \mathbf{G}', \mathbf{q} + \mathbf{G}, \omega)$ and of $\epsilon^{-1}(\mathbf{q} + \mathbf{G}', \mathbf{q} + \mathbf{G}, \omega)$ requires the inversion of a matrix that can easily reach formidable sizes, computationally non-tractable. In the case of crystals, a procedure can be conveniently introduced to overcome this problem. First, the polarizability $\chi(\mathbf{q} + \mathbf{G}', \mathbf{q} + \mathbf{G}, \omega)$ is considered. Second, the Bloch functions are expressed in terms of either Wannier functions or atomic orbitals [24].

The connection between induced charge n_{ind} and total potential is expressed by the relation [25]

$$n_{\text{ind}}(\mathbf{q} + \mathbf{G}', \omega) = \sum_{\mathbf{G}} \chi(\mathbf{q} + \mathbf{G}', \mathbf{q} + \mathbf{G}, \omega) \phi(\mathbf{q} + \mathbf{G}, \omega). \quad (6.139)$$

The dielectric function is given by

$$\varepsilon(\mathbf{q} + \mathbf{G}, \mathbf{q} + \mathbf{G}', \omega) = \delta_{\mathbf{G}, \mathbf{G}'} - v(\mathbf{q} + \mathbf{G}) \chi(\mathbf{q} + \mathbf{G}, \mathbf{q} + \mathbf{G}', \omega), \quad (6.140)$$

with $v(\mathbf{q}) = 4\pi e^2 / (V_c q^2)$ and V_c the volume of the elementary cell in the lattice. In the case of crystals the calculation of polarizability is not too complicated, though it is not here reported. As already discussed, measurements are related to the imaginary part of the response function by means of the fluctuation-dissipation theorem. The quantity that is observed in experiments is in fact

$$\varepsilon(\omega) = \lim_{\mathbf{q} \rightarrow 0} \varepsilon(\mathbf{q}, \mathbf{q}, \omega). \quad (6.141)$$

Fig. 6.10 shows the imaginary part of $\varepsilon(\omega)$ for diamond as calculated within three different approximations: (a) the so-called Random-Phase Approximation (RPA), corresponding to set $f_{xc} = 0$; (b) RPA with inclusion of exchange; and (c) with full inclusion of f_{xc} , where the Bloch functions and energy bands calculations are performed by means of the DFT method. It is seen that only the latter calculation provides a good agreement with experimental data.

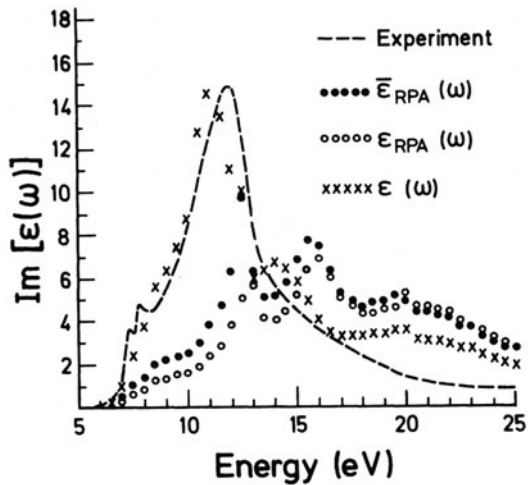


Fig. 6.10 Dielectric properties of crystals. $\text{Im}\varepsilon(\omega)$ for diamond. The comparison between (a) RPA, (b) RPA with exchange and (c) with full inclusion of f_{xc} within the DFT method is shown. Reprinted with permission from [24]. Copyright (1978) by the American Physical Society

6.10.3 Application of DFT to electronic and optical properties in semiconductor nanostructures: the case of a quantum dot

In Chap. 5, the striking consequences of quantum confinement on optical properties of nanostructures has been discussed, stepping on the example behavior of 2D quantum wells and 0D quantum dots. The case of electronic and optical properties of 0D quantum dots is here elaborated, as an application of the DFT tools developed so far. To this aim, it is reminded that in Chap. 2 the problem of electronic states and of excitons in semiconductor nanostructures has been treated within the effective mass approximation. The theoretical predictions have generally shown good agreement with experimental findings. However, especially for very small nanodots, the effective mass approximation is of doubtful reliability. In this case, DFT methods turn out to be useful.

Consider a spherical dot with radius R immersed in a matrix, and a point charge Ze , e.g. a charged impurity, placed at the center of the dot [10]. As compared to the case discussed in Sec. 6.5.3, two lengths are now entering the problem: the screening radius R_s and the dot radius R , with $R_s < R$. The screening radius R_s has already been introduced in Sec. 6.5.3 as a measure of the region size within which an induced charge displacement is observed, after an impurity has been introduced in the material. Let $-Qe$ be the total charge displaced within a sphere of radius R_s . Because the total induced charge must be zero, an amount of charge Qe must be induced between R_s and the dot surface. In the present model, we assume that this charge Qe is uniformly distributed over the dot surface and show that it is responsible for the reduction of screening in a quantum dot. Thus, as referred to the position of charge Ze placed at the origin, three regions are identified where the total potential energy is to be calculated: the regions are such that $0 < r < R_s$, $R_s < r < R$ and $r > R$. In the last two regions, the total potential energy is given by

$$V_c(r) = \begin{cases} -e^2 \left[\frac{Z-Q}{r} + \frac{Q}{R} \right] & \text{for } R_s < r < R, \\ -e^2 \frac{Z}{r} & \text{for } r \geq R. \end{cases} \quad (6.142)$$

Here, the first line expresses the potential energy originated by the charge $e(Z - Q)$ at the sphere center and by the charge eQ induced on the sphere surface. The second line represents the potential energy generated by the charge Ze outside the dot, where all the induced charges compensate with each other. The sphere of induced charge is not in contact with the infinitely extended material, therefore (6.77) is to be modified with a new screening constant k'_0 to be determined. To this aim, the boundary condition (6.66) is used, that has to connect to $V_c(r)$ at $r = R_s$. Since in the region with $R_s < r < R$ no induced charges are present, one has that $\delta\mu = V_c(R_s)$. The solution of (6.77) in the region with $0 < r < R_s$ and satisfying the condition (6.66) in continuity with $V_c(r)$ at $r = R_s$, is given by

$$V(r) = -\frac{Ze^2}{r} \frac{\sinh[k'_0(R_s - r)]}{\sinh(k'_0 R_s)} + V_c(R_s). \quad (6.143)$$

Expressions (6.142) and (6.143) provide the potential energy function in the whole space. On top of the sphere with radius R_s , the electric field has to be continuous as well so that

$$\frac{\sinh(k'_0 R_s)}{k'_0 R_s} = \frac{1}{1 - Q/Z}. \quad (6.144)$$

As already discussed, semiconductors and insulators have finite R_s and thus, one finds that $Q < Z$. In the case of metals with $R_s \rightarrow \infty$, it must occur that $Q = Z$. In any event, the relation (6.144) leads to the determination of k'_0 once Q , R_s and Z are given. One more equation is thus needed, since R_s is unknown. This equation is derived below.

Define now the dielectric constant $\varepsilon(r)$ in real space by means of

$$\varepsilon(r) = \frac{V_e(r)}{V(r)}, \quad (6.145)$$

with $V_e(r) = -Ze^2/r$. One finds

$$\frac{1}{\varepsilon(r)} = \begin{cases} \frac{1 - Q/Z}{k'_0 R_s} [\sinh(k'_0(R_s - r)) + k'_0 r] + \frac{rQ}{RZ} & \text{for } 0 < r < R_s \\ 1 - Q/Z + \frac{rQ}{RZ} & \text{for } R_s < r < R \\ 1 & \text{for } r > R \end{cases} \quad (6.146)$$

The concept of screening sphere is related to the local nature of chemical bonds. Thus one can reasonably assume that R_s be independent of R , so that in the $R \rightarrow \infty$ limit (6.146) converges to the bulk crystal result. One obtains

$$\frac{1}{\varepsilon_{\text{bulk}}(r)} = \begin{cases} \frac{1 - Q/Z}{k'_0 R_s} \{ \sinh[k'_0(R_s - r)] + k'_0 r \} & \text{for } 0 < r < R_s, \\ 1 - \frac{Q}{Z} & \text{for } r > R_s. \end{cases} \quad (6.147)$$

One then finds that in the region with $r > R_s$, the static dielectric constant ε_0 is

$$\varepsilon_0 = \frac{1}{1 - Q/Z}. \quad (6.148)$$

Inserting this relation in (6.147) and performing the Fourier transform, one finds the dielectric function (6.186). In addition, the expressions (6.148) and (6.144) reproduce the equation (6.181), showing that the screening constant k'_0 coincides with k_0 .

The good result is therefore obtained, according to which the finite system imposes a lower dielectric constant with respect to the case of an infinite crystal. However, this result is based on the fact that the electronic kinetic energy has been calculated within the effective mass approximation and the Thomas-Fermi model.

The determination of R_s and k'_0 is now performed by setting the additional required equation. Consider a quasi-spherical dot of Si or Ge: the electronic charge density can be calculated from DFT within LDA, also taking into account atom relaxation processes. Then, an external charge is placed at the dot center, such as P^+ and As^+ , respectively, and the electronic charge distribution is calculated once again: in this manner, charge neutrality is indeed realized by saturating the dangling bonds on the surface with hydrogen atoms. Using the DFT, the induced charge is then calculated by means of

$$n_{\text{ind}} = n [Si_{l-1}PH_m] - n [Si_lH_m], \quad (6.149)$$

where l and m label the number of Si and H atoms, respectively. Once integrated over the sphere with radius R_s , this yields

$$\frac{Q}{Z} = 4\pi \int_0^{R_s} dr r^2 n_{\text{ind}}(r). \quad (6.150)$$

As a consequence, the relation between Q/Z and R_s is obtained. In the cases under consideration, one has $Q = 1$ and the total number of electrons is $N = 4l + m$. The sphere radius R is given by

$$n_0 \frac{4}{3} \pi R^3 = N,$$

with n_0 the bulk density of Si electrons. Equations (6.150) and (6.144) determine R_s and k'_0 . Table 6.2 lists the values of R , R_s and k'_0 in a.u. for selected nanodots of Si and Ge. Notice that in all but the smaller nanodot the values of R_s and k'_0 are

Table 6.2 Electronic and optical properties of nanostructures within DFT. Values of R , R_s and k'_0 (a.u.) for the selected nanodots of Si and Ge listed in the first column

| Nanodot | R | R_s | k'_0 |
|------------------------------------|----------|-------|--------|
| Si ₃₅ H ₃₆ | 11.31 | 5.58 | 0.836 |
| Si ₈₇ H ₇₆ | 15.17 | 5.39 | 0.866 |
| Si ₁₄₇ H ₁₀₀ | 17.82 | 5.35 | 0.873 |
| Si ₁₉₁ H ₁₄₈ | 19.58 | 5.36 | 0.871 |
| Si bulk | ∞ | 5.33 | 0.875 |
| Ge ₁₉₁ H ₁₄₈ | 20.62 | 5.61 | 0.882 |

almost independent on the dot size R : this is a consequence of the local nature of the system response to the external charge. Fig. 6.11 shows the comparison between the integrated induced charge as calculated in nanodots of silicon and germanium,

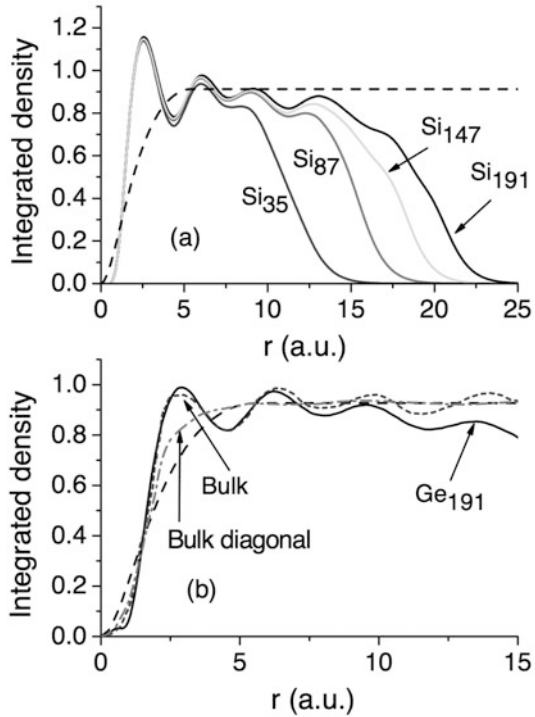


Fig. 6.11 Electronic and optical properties of nanostructures within DFT. Integrated induced charge density profile. (a) Different Si-based nanodots. (b) Ge-based nanodot. Solid lines: DFT calculation. Short-dashed lines: bulk case. Dot-dashed: bulk-diagonal case. Long-dashed lines: Thomas-Fermi model applied to the bulk case. See [10] and references therein

with that obtained in the case of a crystal. Notice that the nanodot charge vanishes at large distance from the origin: this is an artifact of the missing induced surface charge. The space-dependent dielectric function is shown in Fig. 6.12, as calculated by using (6.146) combined with the data in Table 6.2. Screening is seen to be reduced, especially for the smaller dots. Fig. 6.13 shows for $\text{Si}_{35}\text{H}_{36}$ the prediction from (6.146) and the table data for this nanostructure. The prediction is compared with the data referring to the bulk crystal and a full ab initio calculation. In this case, the approximate model yields results very close to those found with the DFT.

In a finite system with $\omega = 0$, the total potential $\phi(\mathbf{r})$ and the total electrostatic energy can be written as

$$\phi(\mathbf{r}_1) = \int d\mathbf{r} \epsilon^{-1}(\mathbf{r}_1, \mathbf{r}) \phi_{\text{ext}}(\mathbf{r}), \quad (6.151)$$

and

$$E_{el} = \int d\mathbf{r}_1 |\psi_e(\mathbf{r}_1)|^2 \phi(\mathbf{r}_1) = \int d\mathbf{r}_1 d\mathbf{r} d\mathbf{r}_2 \frac{1}{\epsilon(\mathbf{r}_1, \mathbf{r})} \frac{|\psi_e(\mathbf{r}_1)|^2 |\psi_h(\mathbf{r}_2)|^2}{|\mathbf{r} - \mathbf{r}_2|}. \quad (6.152)$$

Here, $\psi_e(\mathbf{r}_1)$ and $\psi_h(\mathbf{r}_2)$ are the wavefunctions of the first non-occupied states (LUMO) and of the highest occupied states (HOMO) for the dot. Equation (6.152)

Fig. 6.12 Electronic and optical properties of nanostructures within DFT. Space-dependent dielectric function calculated from (6.146) combined with the data in Table 6.2. Different curves refer to different Si-based nanodots as in the legend [10]

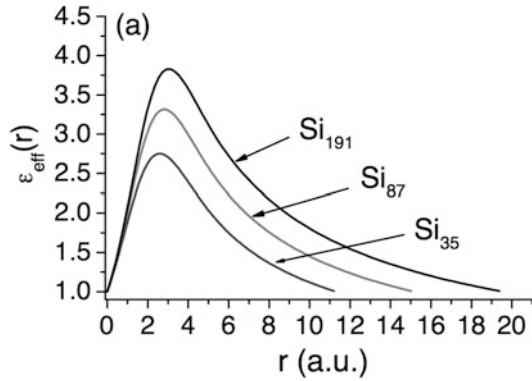
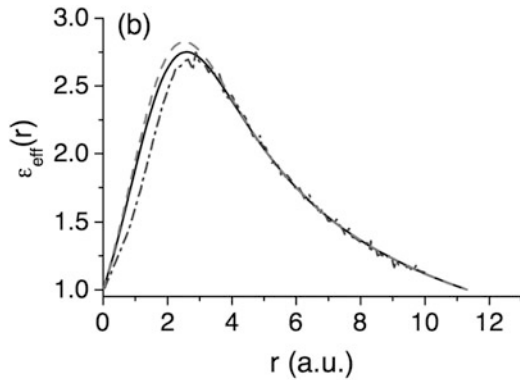


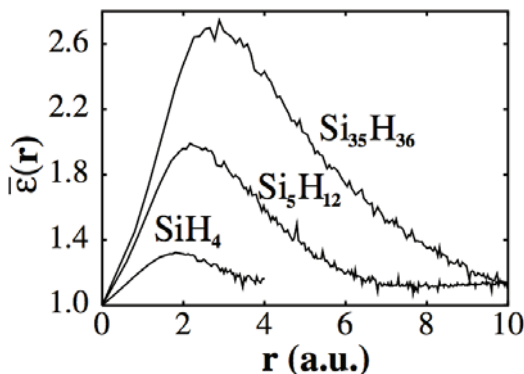
Fig. 6.13 Electronic and optical properties of nanostructures within DFT. Space-dependent dielectric function for the nanodot $\text{Si}_{35}\text{H}_{36}$ in Fig. 6.12. Solid line: results from (6.146) combined with the data in Table 6.2. Dashed line: bulk calculation. Dot-dashed line: ab initio calculation. See [10] and references therein



provides the interaction energy for the exciton. The calculation of E_{el} is computationally very demanding and can be performed only for very small dots by means of the space-dependent dielectric function $\varepsilon(\mathbf{r}, \mathbf{r}')$ [27]. The matrix inversion is especially challenging: the accuracy depends on the number of integration points and on the number of HOMO and LUMO states. The matrix size is the products of all these numbers and turns out to be very large. For this reason, only small dots can be tackled, the largest that has been considered being $\text{Si}_{35}\text{H}_{36}$. Once $\varepsilon^{-1}(\mathbf{r}', \mathbf{r})$ is calculated, one sets $\mathbf{r}' = 0$ and performs the angular average over \mathbf{r} . The quantity $\varepsilon(r) = 4\pi / \int d\Omega \bar{\varepsilon}^{-1}(0, \mathbf{r})$ is plotted in Fig. 6.14: a maximum is observed at intermediate distances r from the center, whereas at each given distance the value of $\varepsilon(r)$ is seen to drop down as long as the size R of the dot radius decreases. Notice the good agreement with the theoretical predictions displayed for $\text{Si}_{35}\text{H}_{36}$ in Fig. 6.13.

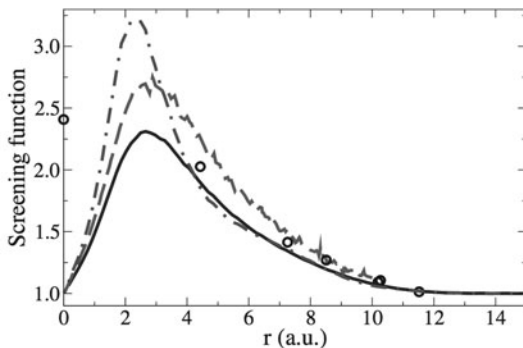
The same procedure used to account for the local field effects in the case of crystals can be extended to nanodots [28]. To this aim, the explicit calculation of $\varepsilon^{-1}(\mathbf{r}', \mathbf{r}, \omega)$ is performed by applying an external perturbation originated by a point-like charge. Fig. 6.15 displays the calculated spherical average $\varepsilon(r, 0) = 4\pi / \int d\Omega \bar{\varepsilon}^{-1}(0, \mathbf{r}, \omega = 0)$ for a $\text{Si}_{35}\text{H}_{36}$ dot. Fig. 6.16 displays $\varepsilon(r, \omega) = 4\pi / \int d\Omega \bar{\varepsilon}^{-1}(0, \mathbf{r}, \omega)$ for a $\text{Si}_{191}\text{H}_{148}$ dot, using static screening with $\omega = 0$ and

Fig. 6.14 Electronic and optical properties of nanostructures within DFT. The quantity $\varepsilon(r) = 4\pi/\int d\Omega \bar{\varepsilon}^{-1}(0, \mathbf{r})$ (see text) is displayed for the different Si-based nanodots as in the legend. Reprinted with permission from [26]. Copyright (2003) by the American Physical Society



optical screening at the gap frequency $\omega_g = E_g/\hbar$. Notice that the two curves show a maximum in correspondence of the same value of r , though the maximum of $\varepsilon(r, \omega_g)$ is higher.

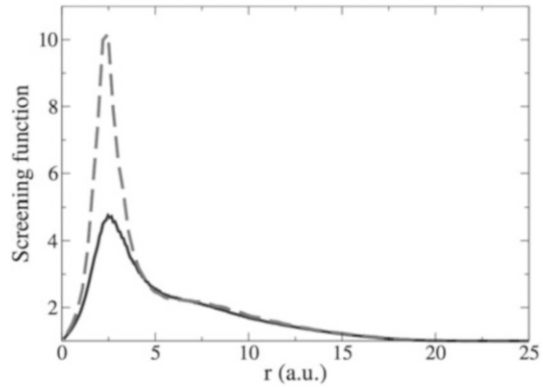
Fig. 6.15 Electronic and optical properties of nanostructures within DFT. spherical average $\varepsilon(r, 0) = 4\pi/\int d\Omega \bar{\varepsilon}^{-1}(0, \mathbf{r}, \omega = 0)$ for a $\text{Si}_{35}\text{H}_{36}$ dot. Solid line and dot-dashed line: calculation described in the text with two different on-site repulsion parameters. Dashed line: ab-initio calculation. Open circles: tight-binding calculation. See [28] and references therein. Reprinted with permission from [28]. Copyright (2006) by the American Physical Society



6.11 Time-Dependent Density Functional Theory and connection with current-density response theory

A conceptual and practical framework has been built up so far, to treat even strongly interacting systems with many particles, in the presence of weak inhomogeneities. Finite temperature and dynamical effects are for now left out. The question arises to whether it is possible to extend the DFT idea to describe the dynamical behavior of a inhomogeneous quantum system at finite temperature in terms of its homogeneous counterpart. It can be anticipated that a framework similar to that of DFT can be set, where however the functionals are to be built up from the current density

Fig. 6.16 Electronic and optical properties of nanostructures within DFT. Spherical average $\varepsilon(r, \omega) = 4\pi / \int d\Omega \varepsilon^{-1}(0, \mathbf{r}, \omega)$ for a $\text{Si}_{191}\text{H}_{148}$ dot. Solid line: static screening with $\omega = 0$. Dashed line: optical screening at the gap frequency $\omega_g = E_g/\hbar$. Reprinted with permission from [28]. Copyright (2006) by the American Physical Society



instead than from the density. This matter is at the same time conceptually simple and complex. It is simple because its bottom line is composed of conservation laws for matter, energy, linear and angular momenta, and principles of invariance under Galileian transformations and time-reversal symmetry operations. It is complex because these essential principles are built in the equations while managing at the same time many particles, possibly strongly interacting, with spatial inhomogeneities and time-dependent behavior: everything at a microscopic, quantum level. However, the conservation laws strictly drive the form that the dynamical equations must have, independently of their inner complexities.

In order to follow this complex matter, the following strategy is thus adopted. The form of the equations of motion is first discussed at a phenomenological level, using the far-running perspective of the conservation laws. These are the Navier-Stokes equations. Then, magnifying glasses are put on, and the correspondence between the macroscopic and microscopic quantities is set up. At the end, all the tiles of the puzzle composed in Chaps. 1-5 result to be connected together and give the whole picture.

6.11.1 Introduction: the Navier-Stokes equations

As seen in frequency domain, the description of the dynamical behavior of an inhomogeneous interacting quantum system at finite temperature in terms of its homogeneous non-interacting counterpart is in fact a crucial question. It is well known that fluids show no resistance to shear at low frequencies, but this is might not be case at high frequency: liquids with large molecules may show macroscopic viscoelastic behavior at moderate flow rates. The frequency spectrum of viscosities is an important piece of knowledge to understand their dynamical behavior.

In the presence of slowly-varying perturbations in space and time, a hydrodynamic approximation can be used in which physical dynamical quantities experience local equilibrium conditions. In each portion of space, local densities of par-

ticles $n(\mathbf{r}, t)$, current $\mathbf{J}(\mathbf{r}, t)$ and energy $\varepsilon(\mathbf{r}, t)$ can be defined, that are conserved on a local scale. This means that each of them satisfies a continuity equation. In order to keep notation easier, from now on the following correspondences are set $n(\mathbf{r}, t) \rightarrow n$, $\mathbf{J}(\mathbf{r}, t) \rightarrow \mathbf{J}$, and $\varepsilon(\mathbf{r}, t) \rightarrow \varepsilon$ to label the values of the densities averaged over the equilibrium ensemble. The continuity equations read

$$\begin{aligned}\frac{\partial n}{\partial t} + \nabla \cdot \mathbf{J} &= 0 \\ \frac{\partial J_i}{\partial t} + \nabla_j \Pi_{ij} &= 0 \\ \frac{\partial \varepsilon}{\partial t} + \nabla \cdot \mathbf{J}_\varepsilon &= 0.\end{aligned}\tag{6.153}$$

The first equation states that matter is conserved: the rapidity with which the system density changes in time within the local volume portion, is to be compensated by the net current flowing in and out of that portion. The second equation simply states that because of linear and angular momentum conservation, the time-derivative of the i -component of the current changes in time, that in essence is a mass times an acceleration along direction i , must be originated by a force. The new concept here is that the i -component of this force is the divergence ∇_j of a symmetric tensor of second rank, that is Π_{ij} : this is called stress tensor. Similarly to the first continuity equation for the density, the last equation simply states that the time derivative of energy is balanced by the divergence of an energy current \mathbf{J}_ε . Notice that all the above processes have been already met earlier in this textbook. The question now remains on how Π_{ij} and \mathbf{J}_ε are shaped. General arguments help to provide the answer.

In the absence of dissipative processes, invariance under Galileian transformations must hold, that is the content of the first Newton law. When implementing this concept to linear order in the velocities, this ties the form of the stress tensor to be Π_{ij} and of the energy current \mathbf{J}_ε . A time change in the current density is restored by a spatial change of pressure: thus, the stress tensor is to represent pressure fluctuation, or $\Pi_{ij} = p\delta_{ij}$. A time change of energy current is restored by either an energy ε - or pressure p -driven spatial change, with velocity $\mathbf{v} = \mathbf{J}/(mn)$ in terms of mass m and density n : thus, $\mathbf{J}_\varepsilon = (\varepsilon + p)\mathbf{v}$. These are examples of restoring, conservative, forces.

Dissipation is now to be included. Dissipative forces acting on the current density are driven by spatial changes $\nabla_j \sigma_{ij}$ of a second stress tensor σ_{ij} , which embodies the viscous behavior: in fact the so-called viscoelastic stress tensor σ_{ij} is built up by linear expansion in the spatial changes of the velocity field $\mathbf{v}(\mathbf{r})$, via bulk ζ and shear η viscosity coefficients. As discussed in Sec. 2.8.3 of Chap. 2, the velocity field $\mathbf{v}(\mathbf{r})$ plays a similar role as the vector potential does in Maxwell equations. Dissipative forces acting on energy changes are driven by gradients of the local temperature $T(\mathbf{r}, t) \rightarrow T$, via the heat conductivity coefficient κ , already met in Chap. 4 when describing transport.

Collecting all these arguments together, to linear order in the driving fields one obtains:

$$\begin{aligned}
\frac{\partial n}{\partial t} + \nabla \cdot \mathbf{J} &= 0 \\
\frac{\partial J_i}{\partial t} + \nabla_j [p\delta_{ij} - \sigma_{ij}] &= 0 \\
\frac{\partial \varepsilon}{\partial t} + \nabla \cdot [(\varepsilon + p)\mathbf{v} - \kappa \nabla T] &= 0.
\end{aligned} \tag{6.154}$$

Notice that the restoring forces $\nabla_j [p\delta_{ij}]$ and $\nabla \cdot [(\varepsilon + p)\mathbf{v}]$ correctly appear with a plus sign on the left-hand-side of (6.154), whereas the dissipative forces $\nabla \cdot [-\kappa \nabla T]$ and $\nabla_j [-\sigma_{ij}]$ with a minus sign.

In (6.154), the detailed form of the viscoelastic stress tensor is:

$$\sigma_{ij} = \tilde{\eta} \left(\frac{\partial v_i}{\partial r_j} + \frac{\partial v_j}{\partial r_i} - \frac{2}{3} \nabla \cdot \mathbf{v} \delta_{ij} \right) + \tilde{\zeta} \nabla \cdot \mathbf{v} \delta_{ij}. \tag{6.155}$$

It includes shear dissipative processes as driven by the rotor of the velocity field via the shear viscosity η , and longitudinal dissipative processes as driven by the divergence of the velocity field via the bulk viscosity ζ .

The first and last equations in (6.154) can be combined together and give the more expressive form

$$\frac{\partial (s\varepsilon)}{\partial t} - \kappa \frac{\nabla^2 T}{T} = 0, \tag{6.156}$$

for the time derivative of the local entropy density $s(\mathbf{r}, t) = \varepsilon(\mathbf{r}, t) - (\varepsilon + p)n(\mathbf{r}, t)/n \rightarrow s$ where now the space and time dependence has been explicitly inserted when needed.

One may ask why no pressure nor temperature gradients like ∇p and ∇T enter (6.154) to drive dissipative currents. While the proof of the statement is not trivial, it can be demonstrated that this is a result of invariance with respect to time-reversal symmetry. Equations (6.154) with the last one possibly replaced by (6.156) are the so-called Navier-Stokes equations, written linear order in the velocities.

Two more tiles of the puzzle are still missing. First, non-dissipative, restoring forces are connected to thermodynamic derivatives. This is a crucial step, since thermodynamic derivatives are measurable macroscopic quantities. Indeed, the restoring force ∇p can be viewed as:

$$\nabla p(\mathbf{r}, t) = \left(\frac{\partial p}{\partial n} \right)_s \nabla n(\mathbf{r}, t) + \frac{V}{T} \left(\frac{\partial p}{\partial s} \right)_n \nabla T s(\mathbf{r}, t),$$

simply stating that a pressure gradient can be originated by either a density or a temperature gradient, in terms of the appropriate thermodynamic derivatives of pressure with respect to density and temperature are included, respectively. For example, the first thermodynamic derivative involved here is $(\partial p / \partial n)_s = mc^2$, with c the speed of sound. More in general, the thermodynamic derivatives are connected to the static susceptibilities introduced in Chap. 4, i.e. to static density response functions in the long-wavelength $k \rightarrow 0$ limit.

Second, as already explored in Chap. 5, dissipated power is related to the imaginary part of the density response function. In fact, Kubo has demonstrated that the dissipative transport coefficients are connected to the imaginary part of the longitudinal and transverse current response functions by means of

$$\begin{aligned}\lim_{\omega \rightarrow 0} \lim_{k \rightarrow 0} \text{Im} \left[\frac{\omega}{k^2} \chi_{jj,L}(k, \omega) \right] &= \frac{4}{3} \eta + \zeta \\ \lim_{\omega \rightarrow 0} \lim_{k \rightarrow 0} \text{Im} \left[\frac{\omega}{k^2} \chi_{jj,T}(k, \omega) \right] &= \eta. \\ \lim_{\omega \rightarrow 0} \lim_{k \rightarrow 0} \text{Im} \left[\frac{\omega}{k^2} \chi_{\varepsilon\varepsilon}(k, \omega) \right] &= \kappa T.\end{aligned}\tag{6.157}$$

Notice that to obtain the dissipative transport coefficients, the $k \rightarrow 0$ limit has to be executed first, followed by the $\omega \rightarrow 0$ limit afterward, whereas the opposite order is to be followed when calculating the static susceptibilities. A pedagogical derivation of Kubo formulas can be found in the work of Martin [30].

6.11.2 Time Dependent-DFT and microscopic Navier-Stokes equations

The Navier-Stokes equations (6.154) have been derived within a phenomenological approach along with the use of general conservation laws and symmetry arguments. The same general arguments can be used within a microscopic formulation, that stems from the Density Functional Theory adapted to include current densities for the description of Time-Dependent processes. The result is anticipated that the form of the microscopic Navier-Stokes equations remains unchanged with respect to the phenomenological (6.154), but the different physical quantities acquire a quantum many-particle microscopic definition. The conceptual and theoretical framework of Time-Dependent DFT (TD-DFT) is sketched below, paralleling the derivations of DFT.

Extension of DFT to the theoretical investigation of time-dependent systems has started with the theorem of Gross and Kohn [29], paralleling that by Hohenberg and Kohn. As to the analogue of the Kohn-Sham scheme, the somehow obvious idea has been pursued of using a modification of Local-Density Approximation extended to take into account conditions of slow variations in time besides than in space. In this so-called Adiabatic Local Density Approximation, the local density to be substituted into the energy functional of the homogeneous system was the time-dependent local density. Unfortunately, ALDA is found to break a number of basic theorems and symmetries. Vignale and Kohn [16] have overcome this difficulties by formulating the TD-DFT in terms of the current density, involving current-current response functions instead than density-density response as in ordinary DFT.

The key idea is that currents are originated by vector potentials. Vignale and Kohn have provided an explicit expression for a linearized exchange and correlation

vector potential $\mathbf{A}_{xc}(\mathbf{r}, \omega)$ in a system with slowly-varying density, subject to a spatially slowly-varying and frequency-dependent external vector potential. The slow variation is to be intended as $k, q \ll \omega/v_f$ and $k, q \ll k_f$, that is the typical length scales k^{-1} and q^{-1} of external potential and equilibrium densities are to be longer than the other length scales dictated by the Fermi wavevector k_f and v_f/ω , in terms of the Fermi velocity v_f . The form of $\mathbf{A}_{xc}(\mathbf{r}, \omega)$ found in [16] reflects and embodies the same Galileian invariance and the zero force and zero torque theorems as discussed for the Navier-Stokes equations: in essence, it is assessed that the system has to move with constant linear and angular velocities if no external forces or torques are applied. A much more expressive form for $\mathbf{A}_{xc}(\mathbf{r}, \omega)$ has then be derived in [18], that leads to an alternative formulation where the equation of motion for the current is formally identical to the linearized hydrodynamic Navier-Stokes equations for a classical fluid and related viscosities [30].

Consider a system at equilibrium, characterized by the equilibrium density $n_0(\mathbf{r})$. Since the equations look much simpler in frequency domain, Fourier transform with respect to time is from now on systematically operated. Apply therefore a weak external vector potential $\mathbf{A}_e(\mathbf{r}, t) = \mathbf{A}_e(\mathbf{r}, \omega)e^{-i\omega t}$. A current density $\mathbf{j}(\mathbf{r}, t) = \mathbf{j}(\mathbf{r}, \omega)e^{-i\omega t}$ is induced to first order in \mathbf{A} , dictated by the current-current response function. Following the Kohn-Sham scheme and its connection with response function theory discussed in Sec. 6.8, the Kohn-Sham current-current response function $\chi_{KS,ij}(\mathbf{r}, \mathbf{r}', \omega)$ of the non-interacting system with equilibrium density $n_0(\mathbf{r})$ can be defined. The induced current is thus

$$j_i(\mathbf{r}, \omega) = \int \sum_j \chi_{KS,ij}(\mathbf{r}, \mathbf{r}', \omega) [A_{ej}(\mathbf{r}', \omega) + A_{H,j}(\mathbf{r}', \omega) + A_{xc,j}(\mathbf{r}', \omega)] d\mathbf{r}'. \quad (6.158)$$

Here, $\mathbf{A}_H(\mathbf{r}, t) = c \int^t \nabla V_H(\mathbf{r}, t') dt'$ is the linearized vector Hartree potential: in the present gauge it is built up from the gradient of the scalar Hartree potential V_H used in DFT, and is indeed longitudinal in character. \mathbf{A}_{xc} is instead the linearized exchange and correlation vector potential which contains all the interaction effects. At variance with \mathbf{A}_H that is always longitudinal, \mathbf{A}_{xc} has in general both longitudinal and transverse components.

Explicit calculation of the current response leads to the following linearized Navier-Stokes equation for the i -component j_i of the current density [30] with complex and frequency-dependent viscosity coefficients:

$$\begin{aligned} -iM\omega j_i(\mathbf{r}, \omega) = n_0(\mathbf{r}) \Big[& -i\omega \frac{e}{c} A_{ei}(\mathbf{r}, \omega) \\ & - \nabla_i \left(\frac{\delta p(n)}{n_0} + V_H(\mathbf{r}, \omega) + V_{xc}^{LDA}(\mathbf{r}, \omega) \right) \Big] \\ & + \sum_j \frac{\partial \sigma_{xc,ij}(\mathbf{r}, \omega)}{\partial r_j}. \end{aligned} \quad (6.159)$$

Notice that (6.159) has the same form as the phenomenological Navier-Stokes equation for the current density in (6.154), and an energy-current equation can in fact be written in a similar manner: in fact, general considerations of symmetry and invariance fix the structure of these equation of motion for the current. Therefore, no further comments are spent on the form of (6.159), but the following. The visco-elastic stress tensor $\sigma_{xc,ij}$

$$\sigma_{xc,ij} = \bar{\eta}_{xc} \left(\frac{\partial v_i}{\partial r_j} + \frac{\partial v_j}{\partial r_i} - \frac{2}{3} \nabla \cdot \mathbf{v} \delta_{ij} \right) + \tilde{\zeta}_{xc} \nabla \cdot \mathbf{v} \delta_{ij}, \quad (6.160)$$

is expressed in terms of complex shear and bulk viscosity coefficients $\eta_{xc}(\omega, n_0(\mathbf{r}))$ and $\zeta_{xc}(\omega, n_0(\mathbf{r}))$, respectively.

The recipe is now needed to calculate these viscosities from the homogeneous system, with insertion of the local density corresponding to the equilibrium system. Heavy mathematical elaborations are needed, which in essence use: the connections between DFT and response function theories as in Sec. 6.8, and the continuity equation $\partial_t n + \nabla \cdot \mathbf{j} = 0$ to connect density and current response functions. In spite of the mathematical complexity, the main results are now given in order to complete the picture. They are expressed in terms of the the excess energy ϵ_{xc} and of the exchange and correlation kernels $f_{xc}^h(\omega, n)$ introduced in Sec. 6.8 and connecting DFT and response-function quantities: these refer to the homogeneous system and enter the dynamical local field factor in similar manner than the static case. At variance with Sec. 6.8 indeed, and following the Vignale and Kohn reasoning, here the kernels are related to current-current response functions, and can be longitudinal $f_{xc,L}^h$ or transverse $f_{xc,T}^h$.

Reverting back to the viscosities then, it can be demonstrated that the bulk viscosity is given by

$$\zeta_{xc}(\omega, n) = -\frac{n^2}{i\omega} \left[f_{xc,L}^h - \frac{4}{3} f_{xc,T}^h - \frac{d^2 \epsilon_{xc}(n)}{dn^2} \right], \quad (6.161)$$

and the shear viscosity by

$$\eta_{xc}(\omega, n) = -\frac{n^2}{i\omega} f_{xc,T}^h. \quad (6.162)$$

Notice that both longitudinal and transverse kernels enter the bulk viscosity, while only the transverse one enters the shear viscosity. In particular, the form $f_{xc,ij}^h(\omega, n) = f_{xc,L}^h(\omega, n) k_i k_j + f_{xc,T}^h(\omega, n) (k^2 \delta_{ij} - k_i k_j)$ explicitly separates the longitudinal L and transverse T parts. In particular,

$$f_{xc,L(T)}^h(\omega, n) \equiv -\lim_{k \rightarrow 0} \mathbf{v}_{\mathbf{k}} G_{L(T)}(k, \omega), \quad (6.163)$$

defining the dynamical local field factors $G_{L(T)}(k, \omega)$ in terms the Fourier transform of the pair potential $\mathbf{v}_{\mathbf{k}}$, as long as it has been performed in Sec. 6.8. Ultimately then, the f_{xc} can be determined by calculating the current response function χ_{ij} in

the homogeneous system by, e.g., theoretical perturbative or Quantum Monte Carlo simulational methods. In particular, in real space one has

$$f_{xc,ij}(\mathbf{r}, \mathbf{r}', \omega) = \frac{1}{\chi_{KS,ij}(\mathbf{r}, \mathbf{r}', \omega)} - \frac{1}{\chi_{ij}(\mathbf{r}, \mathbf{r}', \omega)} - \nabla_i v(|\mathbf{r} - \mathbf{r}'|) \nabla'_j, \quad (6.164)$$

in terms of the pair potential $v(|\mathbf{r} - \mathbf{r}'|)$. In the case of electrons, so that $v(|\mathbf{r} - \mathbf{r}'|) = e^2/|\mathbf{r} - \mathbf{r}'|$, $v_{\mathbf{k}} = 4\pi e^2/k^2$ and in the $k \rightarrow 0$ limit, one finds

$$f_{xc,L(T)}^h(\omega, n) = \frac{\omega^2}{k^2} \left(\frac{1}{\chi_{KS,L(T)}^h(\omega, n) + n/M} - \frac{1}{\chi_{L(T)}^h(\omega, n) + n/M} \right) - v_{\mathbf{k}(L,T)} \quad (6.165)$$

with n the density, $v_{\mathbf{k},L} = v_{\mathbf{k}}$ and $v_{\mathbf{k},T} = 0$. The factors n/M in (6.165) are needed to properly account for the continuity equation $\partial_t n + \nabla \cdot \mathbf{j} = 0$ and the commutation relation $[n(r), j_i(r')] = -i \nabla_i \delta(r - r')$. These indeed imply the following relation between density χ_{nn} and current χ_{jj} response functions: $\chi_{nn} = k^2(\chi_{jj,L} + n/M)/\omega^2$.

Notice that since the local field factors are singular for small k and ω , as anticipated this indeed makes a difference whether either the $k \rightarrow 0$ or the $\omega \rightarrow 0$ limit is performed first. The kernels f_{xc}^h are defined by taking the limit $k \rightarrow 0$ first and provide transport coefficients and are the equivalent of a Kubo relation. When the limit $\omega \rightarrow 0$ is performed first, one gets $\lim_{k \rightarrow 0} 4\pi e^2 G_L(k, 0)/k^2 = -d^2 \epsilon_{xc}(n)/dn^2$: this yields a static susceptibility that is a thermodynamic derivative, in this case a compressibility.

The case of superfluid systems. Quite similar though even more complex analysis can be applied to superfluid systems [19, 20], with a few but crucial distinctions. The analogues of the Navier-Stokes equations are the Landau hydrodynamic equations discussed in Sec. 2.8.3 for the two fluids, including a normal and a superfluid component [7].

In order to catch the main game rules, isothermal conditions are for simplicity assumed: the couplings between density and temperature fluctuations are in the following neglected. Reminding the reasoning in Sec. 2.8.3, two current fields are to be considered in this case: the total current \mathbf{j} and the superfluid velocity \mathbf{v}_s . The superfluid velocity $\mathbf{v}_s = \hbar \nabla \theta / M$ is defined as a gradient of the global phase θ of the coherent superfluid, so that the wavefunction of the macroscopically occupied ground state with density $n(r)$ is $\psi(r) = \sqrt{n(r)} e^{i\theta}$. The gradient ensures that superfluid be irrotational, or $\nabla \wedge \mathbf{v}_s = 0$, meaning that it does not rotate when twisted. Above a critical rotation frequency, formation of a vortex can be energetically favorable for the superfluid, that would be quantized.

The form of Landau's equations for the total current density \mathbf{j} and for the superfluid velocity \mathbf{v}_s is dictated by general considerations about symmetries and conservation laws. As for a normal fluid, invariance under a Galileian transformation must hold, along with time-reversal symmetry, and zero-force and zero-torque theorems ensuring conservation of linear and angular momentum. To this considerations, the peculiar irrotational character of the field \mathbf{v}_s is to be added, at least below the threshold for formation of vortices.

The internal restoring forces are now to be identified. These are due to fluctuations in density of particles and entropy. The dissipative forces instead are determined by the normal-fluid velocity \mathbf{v}_n that acts as a vector potential and by the interdiffusion current $\mathbf{j}_r = \rho_s(\mathbf{v}_s - \mathbf{v}_n)$, with ρ_s being the superfluid density. The restoring forces originate the dispersion of the superfluid excitation modes that, as seen in Appendix 2.13 of Chap. 2, are sound modes. The dissipative forces determine the attenuation of these modes through four bulk and one shear viscosity coefficients. In fact, two being the responding currents, four longitudinal and one transverse response functions are possible, that are schematically χ_{jLjL} , $\chi_{jLv_s} = \chi_{v_sjL}$ by time-reversal symmetry, $\chi_{v_sv_s}$, and χ_{jTjT} .

Thus, the form of the generalized hydrodynamic equations is [19]

$$\begin{aligned} -iM\omega j_i(\mathbf{r}, \omega) &= -\nabla_i[\delta p(\mathbf{r}, \omega)\delta_{ij}] + \nabla_j[\delta\sigma_{ij}(\mathbf{r}, \omega)] \\ -iM\omega v_{s,i}(\mathbf{r}, \omega) &= -\nabla_i \cdot [\delta\mu(\mathbf{r}, \omega)\delta_{ij}] + \nabla_j \cdot [\delta\sigma_{ij}^s(\mathbf{r}, \omega)], \end{aligned} \quad (6.166)$$

where σ , σ^s are second rank symmetric tensors. In particular, one finds that conservation laws and symmetry arguments are satisfied by the following expressions for the viscoelastic stress tensors:

$$\begin{aligned} \sigma_{ij} &= \left(\eta(\omega) - \frac{p_0(n)}{i\omega} \right) \left(\frac{\partial v_{ni}}{\partial r_j} + \frac{\partial v_{nj}}{\partial r_i} - \frac{2}{3} \nabla \cdot \mathbf{v}_n \delta_{ij} \right) + \delta_{ij} [\zeta_2(\omega) \nabla \cdot \mathbf{v}_n + \zeta_1(\omega) \nabla \cdot \mathbf{j}_r] \\ \sigma_{ij}^s &= \delta_{ij} [\zeta_3(\omega) \nabla \cdot \mathbf{j}_r + \zeta_4(\omega) \nabla \cdot \mathbf{v}_n]. \end{aligned} \quad (6.167)$$

The xc subscript to the viscosities has been omitted for simplicity. Furthermore, δp and $\delta\mu$ represent the local pressure and chemical potential fluctuations. They are related to fluctuations of the density δn and entropy density δs by means of the thermodynamic derivatives $\delta p(\mathbf{r}, \omega) = (nK_T)^{-1} \delta n(\mathbf{r}, \omega)$ and $\delta\mu(\mathbf{r}, \omega) = (nK_T)^{-1} \delta n(\mathbf{r}, \omega) + (Ts/c_V)^{-1} \delta s(\mathbf{r}, \omega)$, related to the following static susceptibilities: the isothermal compressibility K_T , entropy s and specific heat per particle at constant volume c_V .

Once again, explicit calculation of the current response function with the use of the conservation laws and the symmetries introduced above, leads to microscopic Landau two-fluid equations formally identical to (6.166), where all the various quantities have a microscopic definition. Besides the currents and dissipative forces already introduced, the superfluid density turns out to be a functional derivative of the total current density with respect to the superfluid velocity, and is therefore a response function: as already discussed in the phenomenological treatment of Chap. 2, it essentially measures the response of the system to a twist. The viscosities are related to the new excess kernels f_{xc} , which in turn are related to the longitudinal and transverse current-current response functions of the fluid via the Kubo formulae [30]:

$$\begin{aligned} \lim_{\omega \rightarrow 0} \lim_{k \rightarrow 0} \text{Im} \left[\frac{\omega}{k^2} \chi_{jj,L}(k, \omega) \right] &= \frac{4}{3} \eta + \zeta_2 \\ \lim_{\omega \rightarrow 0} \lim_{k \rightarrow 0} \text{Im} \left[\frac{\omega}{k^2} \chi_{jv_s}(k, \omega) \right] &= \zeta_1 = \zeta_4 \end{aligned}$$

$$\begin{aligned}\lim_{\omega \rightarrow 0} \lim_{k \rightarrow 0} \text{Im} \left[\frac{\omega}{k^2} \chi_{\mathbf{v}_s \mathbf{v}_s}(k, \omega) \right] &= \zeta_3 \\ \lim_{\omega \rightarrow 0} \lim_{k \rightarrow 0} \text{Im} \left[\frac{\omega}{k^2} \chi_{\mathbf{j}\mathbf{j}, T}(k, \omega) \right] &= \eta.\end{aligned}\quad (6.168)$$

Use of additional identities coming from conservation laws leads to the demonstration that the viscosities are to be calculated in the homogeneous system, as for example by means of theoretical or simulational methods, at the local density. Obviously, the Landau hydrodynamic equations are recovered in the appropriate limit, that is homogeneous superfluid under low-frequency drives [19, 20].

In essence

Concept

A unified theoretical framework exists to treat the dynamics of weakly inhomogeneous normal fluids as well as superfluids, that is derived from a formulation of DFT in terms of currents and with the use of general considerations such as Galileian invariance, conservation laws and time-reversal symmetries. Explicit calculation of the current response leads to equations of motion for the currents that are formally equivalent to phenomenological Navier-Stokes equations for a normal fluid and to Landau two-fluids equations for superfluids. In fact, they are formulated in terms of the following microscopic ingredients: (i) microscopic expression for the equilibrium normal and superfluid densities and normal and superfluid currents, the superfluid density being in fact a measure of the system response to an external twist; (ii) knowledge of restoring and dissipative forces, and (iii) visco-elastic spectra of the homogeneous fluid. The latter are determined from current response and Kubo relations and then evaluated at the local equilibrium density in accord with the TD-DFT concepts and schemes.

Summary: concepts, tools, and procedures to know

Concepts and Tools

- **In quantum world statistics automatically introduces a correlation due to exchange. To this, a correlation originated by the interaction between the particles is to be added.**
- Exchange has negligible effects whenever the single-particle wavefunctions are not overlapped and is quite effective when the particle wavefunctions are delocalized, as in crystals and especially in metals.
- The quantity $n_g(\mathbf{r}, \mathbf{r}')$ can be viewed as the particle density which would be observed when sitting on the particle at position \mathbf{r} .
- The density profile $n_{xc}(\mathbf{r}, \mathbf{r}')$ can be viewed as a hole that is dug in by exchange and correlation processes between the system particles, leading to screening of the interactions.

- The adimensional parameter $r_s = r_0/a_b$ measuring the system density can be identified also with the system coupling strength.
- The excess energy due to exchange and correlations modifies both potential and kinetic energy. This is a purely quantum effect, whatever the statistics of the particles might be.
- Inclusion of the exchange term works to keep the electrons far apart because of the Pauli exclusion principle and therefore lowers the energy. This effect is expected to become progressively negligible while the density lowers.
- In metals spatial correlation effects are relevant and are to be treated to some extent together with the spin correlations.
- Exchange and correlation effects are connected to screening. The quantity ϵ_0 is a measure of the number N_i of induced charges contained within the screening sphere with radius R_s . In an insulator $N_i < Z$, whereas in metal $N_i = Z$ implying $\epsilon_0 \rightarrow \infty$ and $R_s \rightarrow \infty$.
- On quite general grounds, Hohenberg and Kohn theorem states that a one-to-one correspondence exists between the density $n(\mathbf{r})$ of an interacting system and the external potential $V_e(\mathbf{r})$ acting on it.
- According to the Kohn and Sham scheme, the interacting system can be mapped onto an equivalent and effective non-interacting one, characterized by an effective single-particle potential.
- DFT and response or dielectric function theories are strictly connected: DFT effective and xc potentials calculated at the local density are related to the response function of the real system and to a Kohn and Sham response function built up from fictitious single-particle orbitals.
- $V_{xc}(\mathbf{r})$ is a good approximation for the xc energies and the DFT results are in general more accurate and realistic than those obtained within Hartree-Fock calculations.
- A unified theoretical framework exists to treat the dynamics of a weakly inhomogeneous normal and super-fluids, that is derived from a formulation of DFT in terms of currents and with the use of general considerations such as Galileian invariance, conservation laws and time-reversal symmetries. Explicit calculation of the microscopic current response in the homogeneous system leads to equations of motion for the currents that are formally equivalent to Navier-Stokes equations for a normal fluid and to Landau two-fluids equations for superfluids.

Procedures

- Local approximation to the non-local Hartree-Fock exchange potential.
 - Self-consistent solution of the Hartree equation.
 - Calculation of the screening properties within the Thomas-Fermi model.
-

6.12 Appendix. Virial theorem

This theorem can be proved in both classical and quantum frameworks.

6.12.0.1 Classical proof

Consider a particle subjected to a force \mathbf{F} . Given e.g. the x -component of the position, one has

$$\frac{dx^2}{dt} = 2x \frac{dx}{dt},$$

and

$$\frac{d^2x^2}{dt^2} = 2 \left(\frac{dx}{dt} \right)^2 + 2x \frac{d^2x}{dt^2}.$$

Considering the other two components as well, along with the Newton equation one has

$$\frac{1}{2}mv^2 = \frac{1}{4}m \frac{d^2r^2}{dt^2} - \frac{1}{2}\mathbf{F} \cdot \mathbf{r}. \quad (6.169)$$

Assume that the particle is part of a system inside a finite volume. Sum now over all the particles and perform an average over time T . One has

$$\frac{1}{T} \int_0^T dt \sum_i \frac{d^2r_i^2}{dt^2} = \frac{1}{T} \left[\sum_i \frac{dr_i^2}{dt} \right]_0^T \simeq 0,$$

due to the finite volume size and to the fact that the considered quantity has about the same value at $t = 0$ and $t = T$. Therefore one obtains:

$$\langle E_K \rangle = -\frac{1}{2} \langle \sum_i \mathbf{F}_i \cdot \mathbf{r}_i \rangle. \quad (6.170)$$

This is the classical form of the Virial theorem. If the force \mathbf{F}_i acting on particle is decomposed into that on the particle i and due to the interactions between the particles and that exerted from the volume walls, one finds

$$-\frac{1}{2}P \int_S \mathbf{r} \cdot d\mathbf{S} = -\frac{1}{2}P \int_V \nabla \cdot \mathbf{r} dV = -\frac{3}{2}PV,$$

where P is the pressure exerted by the system on the container walls. Equation (6.170) can now be cast in the form

$$3PV = 2\langle E_K \rangle - \langle \sum_i \nabla_{\mathbf{r}_i} \Phi \cdot \mathbf{r}_i \rangle, \quad (6.171)$$

with $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} \Phi$ and Φ the potential energy of the interacting particles.

6.12.0.2 Quantum proof

The quantum demonstration proceeds from the Hamiltonian of N interacting particles, which an external pressure is acting on:

$$H = \sum_i \frac{p_i^2}{2m_i} + V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i) = E_K + V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N).$$

Here, $V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i)$ embodies all the interaction terms. The commutator

$$\left[\sum_j (\mathbf{r}_j \cdot \mathbf{p}_j + \mathbf{p}_j \cdot \mathbf{r}_j), H \right] = 2i\hbar E_K - i\hbar \sum_j \mathbf{r}_j \cdot \nabla_{\mathbf{r}_j} V \quad (6.172)$$

leads to the following expectation values over a stationary state:

$$2\langle E_K \rangle = \langle \sum_j \mathbf{r}_j \cdot \nabla_{\mathbf{r}_j} V \rangle, \quad (6.173)$$

where $\langle [\sum_j (\mathbf{r}_j \cdot \mathbf{p}_j + \mathbf{p}_j \cdot \mathbf{r}_j), H] \rangle = 0$. Decomposing the forces from the walls from those due to internal interactions as in the classical proof, one obtains once again (6.171). The case in which $\Phi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ is a homogeneous function of order α , that is $\Phi \propto r_j^\alpha$ for each j is especially interesting, since it comprises Coulomb-like and dipolar-like pair interactions. In this special case, one obtains

$$\langle \sum_j \mathbf{r}_j \cdot \nabla_{\mathbf{r}_j} \Phi \rangle = \alpha \langle \Phi \rangle. \quad (6.174)$$

6.13 Appendix. Screening effects in insulators within the Thomas-Fermi model with appropriate boundary conditions

In this appendix, the detailed calculation is carried out to determine the screening effects in insulators within the Thomas-Fermi model and appropriate boundary conditions. In insulators, electron polarization as expressed by means of the density fluctuation $n(r) - n_0$, is limited to a sphere with radius R_s centered on the charge eZ , that is called screening sphere: one expects that R_s be of the order of a bonding distance. In metals instead, because of delocalization the screening sphere has an infinitely large radius R_s , as already found in Chap. 5. In the case of insulators, the following condition holds:

$$n(r) - n_0 = \begin{cases} mk_f / (\pi^2 \hbar^2) (\delta\mu - V(r)) & \text{for } r < R_s \\ 0 & \text{for } r \geq R_s \end{cases}. \quad (6.175)$$

Since on the sphere $n(r) - n_0 = 0$, it must be

$$\delta\mu = V(R_s), \quad (6.176)$$

that fixes the chemical potential fluctuation. Since no displaced charge lies outside the sphere, in this region of space with $r > R_s$ the potential is

$$V(r) = -\frac{Ze^2}{\epsilon_0 r}, \quad \text{for } r > R_s, \quad (6.177)$$

with ϵ_0 accounting for the screening produced by the induced charge onto the perturbing charge Ze . Taking into account (6.177) and (6.176), the condition on the potential (6.66) leads to one first boundary condition (6.78)

$$A + B = Ze^2. \quad (6.178)$$

The continuity of the potential at $r = R_s$ provides the second required condition

$$Ae^{-k_0 R_s} + Be^{k_0 R_s} = 0. \quad (6.179)$$

Solving the two equations to obtain A and B , one has

$$V(r) = \begin{cases} -(Ze^2/r)\sinh[k_0(R_s - r)]/\sinh[k_0 R_s] + \delta\mu, & \text{for } r < R_s \\ -Ze^2/(\epsilon_0 r), & \text{for } r \geq R_s \end{cases} \quad \text{with } \delta\mu = -\frac{Ze^2}{\epsilon_0 R_s}. \quad (6.180)$$

A third condition is needed in order to determine the screening sphere radius. This is the continuity condition for the electric field at $r = R_s$, leading to

$$\frac{\sinh[k_0 R_s]}{k_0 R_s} = \epsilon_0 \quad (6.181)$$

Result (6.180) embodies a few interesting consequences. First, the number of induced charges within the screening sphere is

$$N_i = 4\pi \int_0^{R_s} [n(r) - n_0] r^2 dr,$$

or else, by means of (6.175):

$$N_i = \frac{4mk_f}{\pi\hbar^2} \int_0^{R_s} [\delta\mu - V(r)] r^2 dr. \quad (6.182)$$

Substituting (6.180) into (6.182) one obtains

$$N_i = Z \left(1 - \frac{1}{\epsilon_0} \right). \quad (6.183)$$

In essence, ϵ_0 is a measure of the number N_i of induced charges contained within the screening sphere with radius R_s . In an insulator $N_i < Z$, whereas in metal $N_i = Z$ as shown in Chap. 5, implying $\epsilon_0 \rightarrow \infty$ in (6.183) and $R_s \rightarrow \infty$.

The induced charge $\rho_{\text{ind}}(r) = -e(n(r) - n_0)$ is connected to the dielectric function as in (6.64), so that it can immediately be obtained after using (6.175) and (6.180):

$$\rho_{\text{ind}}(r) = -\frac{eZk_0^2}{4\pi r} \frac{\sinh[k_0(R_s - r)]}{\sinh[k_0R_s]}, \quad (6.184)$$

whose Fourier transform is

$$\rho_s(q) = Ze \frac{k_0^2}{k_0^2 + q^2} \left[\frac{\sin[qR_s]}{\epsilon_0 q R_s} - 1 \right]. \quad (6.185)$$

Using the definition (6.62), the dielectric function is finally obtained, that is

$$\epsilon(q) = \frac{k_0^2 + q^2}{\frac{k_0^2}{\epsilon_0} \frac{\sin[qR_s]}{qR_s} + q^2}. \quad (6.186)$$

One immediately sees that $\epsilon(0) = \epsilon_0$, and therefore $\epsilon(0)$ has a finite value in insulators. In a metal instead, (6.186) gets the form

$$\epsilon(q) = \frac{k_0^2 + q^2}{q^2}, \quad (6.187)$$

that diverges while $q \rightarrow 0$. Starting from this equation, one finds that in metals the total potential energy is

$$V(r) = -\frac{Ze^2}{r} e^{-k_0 r}, \quad (6.188)$$

as already found in Chap. 5.

Problems with solutions

6.1. Determine the relationship between the linear and planar electronic density, and the Fermi momentum and energy.

Solution. In the linear case at $T = 0$ one has

$$\frac{L}{2\pi} 2 \int_0^{k_f} dk = N,$$

where L and N are the system size and total number of electrons, respectively. One finds $k_f = n\pi$, with n the linear electronic density. The Fermi energy is given by $\epsilon_f = \hbar^2 k_f^2 / (2m) = \hbar^2 n^2 \pi^2 / (2m)$. The total kinetic energy is

$$E = \frac{L}{2\pi} 2 \int_0^{k_f} \frac{\hbar^2}{2m} k^2 dk = \frac{L}{\pi} \frac{\hbar^2}{2m} \frac{k_f^3}{3} = \frac{L}{\pi} \frac{k_f}{3} \varepsilon_f = \frac{N}{3} \varepsilon_f, \quad (6.189)$$

so that the average kinetic energy per particle is $\bar{\varepsilon} = E/N = \varepsilon_f/3$.

In the planar case one has

$$\frac{S}{(2\pi)^2} 2 \int_0^{k_f} 2\pi k dk = N,$$

with S the surface system area. One finds $k_f = \sqrt{2\pi n}$, with n the planar electronic density. The Fermi energy is $\varepsilon_f = \hbar^2 k_f^2 / (2m) = \hbar^2 2\pi n / (2m)$. The total kinetic energy is

$$\begin{aligned} E &= \frac{S}{(2\pi)^2} 2 \int_0^{k_f} \frac{\hbar^2}{2m} k^2 2\pi k dk = \frac{S}{2\pi} \frac{\hbar^2}{2m} \frac{k_f^4}{2} \\ &= \frac{S}{2\pi} k_f^2 \frac{\varepsilon_f}{2} = S n \frac{\varepsilon_f}{2}, \end{aligned} \quad (6.190)$$

so that the average kinetic energy per particle is $\bar{\varepsilon} = E/N = \varepsilon_f/2$.

6.2. Consider the case of a single particle constrained on a side with length L . The average kinetic energy per particle is $\bar{\varepsilon} = \varepsilon_f/3$, with $\varepsilon_f = \hbar^2 k_f^2 / (2m) = \hbar^2 n^2 \pi^2 / (2m)$. Demonstrate that the kinetic energy functional is

$$E_c = \frac{\hbar^2}{2m} \frac{\pi^2}{3} \int_{-\infty}^{\infty} dx n^3(x), \quad (6.191)$$

and calculate it in the case with $n(x) = (2/L) \sin^2(\pi x/L)$ in the interval $(0, L)$. Calculate then the exact value of the kinetic energy when the particle is in its ground state.

Solution. The exact value of the kinetic energy corresponding to the ground state is $E_{\text{ext}} = \hbar^2 \pi^2 / (2mL^2)$. From (6.189) one finds that the average energy per particle is $E_p = \hbar^2 n^2 \pi^2 / (6m)$. In an interval dx are $n(x)dx$ particles and therefore the total kinetic energy is given by (6.191). Using this result for the present case, one finds

$$E_K = \frac{\hbar^2}{2m} \frac{\pi^2}{3} \left(\frac{2}{L} \right)^3 \int_0^L dx \sin^6 \left(\frac{\pi x}{L} \right).$$

Executing the calculation one finds

$$E_K = \frac{\hbar^2}{2m} \frac{\pi^2}{3} \frac{5}{L^2},$$

that amounts to $8.225/L^2$, except for the factor $\hbar^2/(2m)$. The exact value is $9.870/L^2$, thus a 16.7% larger.

6.3. Consider two particles, the first in the ground state and the second in the first excited state. Calculate the exact kinetic energy of the system, along with its approximated value obtained by means of (6.191). To this aim, take $n(x)$ to be

$$n(x) = \frac{2}{L} \left[\sin^2 \left(\frac{\pi x}{L} \right) + \sin^2 \left(\frac{2\pi x}{L} \right) \right].$$

Solution. The exact kinetic energy is $E_{\text{ext}} = \hbar^2 5\pi^2 / (2mL^2)$. Except for the factor $\hbar^2 / (2m)$, it amounts to $5\pi^2 / L^2 = 49.35 / L^2$. The calculation of E_K is lengthy but simple, leading to the value $14\pi^2 / (3L^2) = 46.06 / L^2$. This corresponds to a 6.67% discrepancy with respect to the exact value.

6.4. Use the variational method to show that the normalized wavefunction $\psi(x) = Ae^{-\alpha x^2}$ with $A = (2\alpha/\pi)^{1/4}$ provides the exact ground-state energy $\hbar\omega/2$ of the harmonic oscillator. Write also the expression for $n(x)$.

Solution. The kinetic and potential energies amount respectively to

$$\begin{aligned} E_K &= \frac{\hbar^2}{2m} \int_{-\infty}^{\infty} dx \left(\frac{d\psi(x)}{dx} \right)^2 = \frac{\hbar^2}{2m} A^2 \int_{-\infty}^{\infty} dx 4\alpha^2 x^2 e^{-2\alpha x^2} = \frac{\hbar^2}{2m} \alpha, \\ E_p &= \frac{1}{2} k \int_{-\infty}^{\infty} dx x^2 \psi^2(x) = \frac{1}{2} k A^2 \int_{-\infty}^{\infty} dx x^2 e^{-2\alpha x^2} = \frac{k}{8\alpha}, \end{aligned} \quad (6.192)$$

where k is the restoring force stiffness. The total energy is

$$E = \frac{\hbar^2}{2m} \alpha + \frac{k}{8\alpha}.$$

This is minimized by $\alpha = \sqrt{km} / (2\hbar)$. Substituting this value inside the expression for E one finds $E_m = \hbar\omega/2$, with $\omega = \sqrt{k/m}$. The function $n(x)$ therefore is

$$n(x) = \psi^2(x) = \left(\frac{2\alpha}{\pi} \right)^{1/2} e^{-2\alpha x^2}. \quad (6.193)$$

6.5. Using (6.191) for the kinetic energy, the expression

$$E_p = \frac{1}{2} k \int_{-\infty}^{\infty} dx x^2 n(x)$$

for the potential energy and expression (6.193) for the density, determine the value of α that minimizes the energy.

Solution. The total energy is given by

$$E = \frac{\hbar^2}{2m} \frac{\pi^2}{3} \int_{-\infty}^{\infty} dx n^3(x) + \frac{k}{2} \int_{-\infty}^{\infty} dx x^2 n(x),$$

and substituting (6.193) in it one finds

$$E = \frac{\hbar^2}{2m} \frac{\pi}{3\sqrt{3}} \alpha + \frac{k}{8\alpha}.$$

Energy minimization yields

$$\alpha = \frac{\sqrt{km}}{\hbar} \left(\frac{3\sqrt{3}}{8\pi} \right)^{1/2} = 0.45 \frac{\sqrt{km}}{\hbar}.$$

The minimum is at least

$$E_m = \frac{1}{\sqrt{2}} \sqrt{\frac{\pi}{3\sqrt{3}}} \hbar \omega = 0.55 \hbar \omega.$$

One sees that E_m differ from the exact result to within a 10% .

6.6. Given the two-electron Hamiltonian

$$H = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} \right) - Ze^2 (\delta(x_1) + \delta(x_2)),$$

calculate the ground-state energy assuming that the spins of the two electrons form a singlet state. Adding then the repulsion term $e^2 \delta(x_1 - x_2)$, perform a perturbative calculation of the two-electron system.

Solution. Consider the eigenvalue problem

$$H_0 \psi(x) = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} - Ze^2 \delta(x) \right) \psi(x) = \varepsilon \psi(x).$$

One finds that the eigenfunction satisfying the continuity condition at the origin is $\psi(x) = Ae^{-\alpha|x|}$, with $A = \sqrt{\alpha}$ the normalization factor. The eigenfunction must satisfy also the boundary condition

$$-\frac{\hbar^2}{2m} (\psi'(0^+) - \psi'(0^-)) - Ze^2 \psi(0) = 0,$$

that is $\alpha = Ze^2 m / \hbar^2$. The eigenvalue ε corresponding to the ground state amounts to

$$\varepsilon_0 = \frac{\hbar^2}{2m} \int_{-\infty}^{\infty} dx \left(\frac{d\psi(x)}{dx} \right)^2 - Ze^2 \psi^2(0),$$

that is

$$\varepsilon_0 = -\frac{(Ze^2)^2}{2} \frac{m}{\hbar^2} = -\frac{Z^2 e^2}{2a_b},$$

with a_b the Bohr radius. Considering the problem of two particles with antisymmetric spin function, the wavefunction of the unperturbed state is

$$\psi(x_1, x_2) = A^2 e^{-\alpha|x_1|} e^{-\alpha|x_2|}. \quad (6.194)$$

To the first order in perturbation expansion, the energy is

$$\begin{aligned} \Delta \varepsilon_0 &= A^4 e^2 \int dx_1 dx_2 e^{-2\alpha|x_1|} e^{-2\alpha|x_2|} \delta(x_1 - x_2) \\ &= A^4 e^2 \int_{-\infty}^{\infty} dx e^{-4\alpha|x|} = \frac{Ze^4}{2} \frac{m}{\hbar^2} = \frac{Ze^2}{2a_b}. \end{aligned} \quad (6.195)$$

In the case with $Z = 2$, the total energy is $\varepsilon = -3e^2/a_b$.

6.7. Perform the variational determination of the ground-state energy of the Hamiltonian

$$H = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} \right) - Ze^2 (\delta(x_1) + \delta(x_2)) + e^2 \delta(x_1 - x_2). \quad (6.196)$$

To this aim, use (6.194) as trial wavefunction.

Solution. Since $A = \sqrt{\alpha}$, one finds for the energy

$$E = \frac{\hbar^2}{m} \alpha^2 - 2Ze^2 \alpha + \frac{e^2}{2} \alpha.$$

Thus, imposing the condition $dE/d\alpha = 0$, one has

$$\alpha = \frac{1}{a_b} \left(Z - \frac{1}{4} \right).$$

The energy, calculated at the found α value yields

$$E_m = -\frac{e^2}{a_b} \left(Z - \frac{1}{4} \right)^2 = -3.06 \frac{e^2}{a_b}.$$

The value is a 2% smaller than the perturbative result.

6.8. Using the form (6.191) for the kinetic energy, write the total energy of the system described by the Hamiltonian (6.196) as a function of electron density $n(x)$. Then, given $n(x) = n_0 e^{-\gamma|x|}$ so that $\int_{-\infty}^{\infty} dx n(x) = 1$, determine the parameter values n_0 and γ that minimize the energy.

Solution. First of all one has that

$$\int_{-\infty}^{\infty} dx n(x) = 2n_0 \int_0^{\infty} dx e^{-\gamma x} = 1,$$

that is $\gamma = 2n_0$. The total energy is thus

$$E = \frac{\hbar^2}{2m} \frac{\pi^2}{3} n_0^3 \int_{-\infty}^{\infty} dx e^{-3\gamma|x|} - 2Ze^2 n_0 + e^2 \int_{-\infty}^{\infty} dx n_0^2 e^{-2\gamma|x|},$$

or else

$$E = \frac{\hbar^2}{m} \frac{\pi^2}{9} \frac{n_0^2}{2} - 2e^2 \left(Z - \frac{1}{4} \right) n_0.$$

The minimization condition $dE/dn_0 = 0$ yields

$$\frac{\hbar^2}{m} \frac{\pi^2}{9} n_0 - 2e^2 \left(Z - \frac{1}{4} \right) = 0,$$

that is

$$n_0 = \frac{1}{a_b} \frac{18}{\pi^2} \left(Z - \frac{1}{4} \right).$$

One then obtains for E_m

$$\begin{aligned} E_m &= \frac{e^2}{a_b} \frac{18}{\pi^2} \left[\left(Z - \frac{1}{4} \right)^2 - 2 \left(Z - \frac{1}{4} \right)^2 \right] \\ &= -\frac{e^2}{a_b} \frac{18}{\pi^2} \left(Z - \frac{1}{4} \right)^2. \end{aligned} \quad (6.197)$$

6.9. Consider only one particle with kinetic energy

$$E_K = \frac{\hbar^2}{2m} \int d\mathbf{r} |\nabla \psi(\mathbf{r})|^2.$$

Show that the latter can also be cast in the form

$$E_K = \frac{\hbar^2}{2m} \frac{1}{4} \int d\mathbf{r} \frac{|\nabla n(\mathbf{r})|^2}{n(\mathbf{r})},$$

where $\psi(\mathbf{r})$ is the wavefunction and $n(\mathbf{r}) = \psi^*(\mathbf{r})\psi(\mathbf{r})$.

Solution. One finds that

$$\begin{aligned} \frac{|\nabla(\psi^*(\mathbf{r})\psi(\mathbf{r}))|^2}{|\psi(\mathbf{r})|^2} &= \frac{1}{|\psi(\mathbf{r})|^2} |\psi^*(\mathbf{r})\nabla\psi(\mathbf{r}) + \psi(\mathbf{r})\nabla\psi^*(\mathbf{r})|^2 \\ &= \frac{4}{|\psi(\mathbf{r})|^2} (|\psi(\mathbf{r})|^2 |\nabla\psi(\mathbf{r})|^2) = 4|\nabla\psi(\mathbf{r})|^2. \end{aligned} \quad (6.198)$$

6.10. Calculate the variation of E_K with respect to changes in $n(x)$. For simplicity, perform the calculation in one dimension.

Solution. To within a factor $\hbar^2/2m$, one has

$$\delta E_K = \frac{1}{4} \int dx \left[\frac{(n'(x) + \delta n'(x))^2}{n(x) + \delta n(x)} - \frac{n'^2(x)}{n(x)} \right]$$

$$\begin{aligned}
&= \frac{1}{4} \int dx \left[\frac{n'^2(x) + 2n'(x)\delta n'(x)}{n(x) \left(1 + \frac{\delta n(x)}{n(x)}\right)} - \frac{n'^2(x)}{n(x)} \right] \\
&= \frac{1}{4} \int dx \frac{1}{n(x)} \left[(n'^2(x) + 2n'(x)\delta n'(x)) \left(1 - \frac{\delta n(x)}{n(x)}\right) - n'^2(x) \right] \\
&= \frac{1}{4} \int dx \frac{1}{n(x)} \left[2n'(x)\delta n'(x) - \frac{n'^2(x)}{n(x)}\delta n(x) \right]. \quad (6.199)
\end{aligned}$$

Integrating by part the first term on the right-hand side of the last line and assuming that the contributions from the boundaries do vanish, one has

$$\begin{aligned}
\int dx \frac{1}{n(x)} 2n'(x) \frac{d\delta n(x)}{dx} &= - \int dx \delta n(x) \frac{d}{dx} \left(\frac{2n'(x)}{n(x)} \right) \\
&= - \int dx \delta n(x) \left(\frac{2n''(x)}{n(x)} - \frac{2n'^2(x)}{n^2(x)} \right). \quad (6.200)
\end{aligned}$$

One therefore obtains for δE_K :

$$\delta E_K = \int dx \left(-\frac{1}{2} \frac{n''(x)}{n(x)} + \frac{1}{4} \frac{n'^2(x)}{n^2(x)} \right) \delta n(x). \quad (6.201)$$

6.11. Consider one particle in a potential $V(x)$. Its total energy is

$$E = \frac{\hbar^2}{2m} \frac{1}{4} \int dx \frac{|n'(x)|^2}{n(x)} + \int dx V(x) n(x).$$

Write down the Euler equations for the density.

Solution. Considering the result of Problem 6.10, one finds δE :

$$\delta E = \int dx \delta n(x) \left[\frac{\hbar^2}{2m} \left(-\frac{1}{2} \frac{n''(x)}{n(x)} + \frac{1}{4} \frac{n'^2(x)}{n^2(x)} \right) + V(x) - \mu \right],$$

with μ the Lagrange multiplier introduced by the condition $\int n(x) dx = 1$. The Euler equation is thus

$$\frac{\hbar^2}{2m} \left(-\frac{1}{2} \frac{n''(x)}{n(x)} + \frac{1}{4} \frac{n'^2(x)}{n^2(x)} \right) + V(x) - \mu = 0,$$

or else

$$-\frac{1}{2} \frac{n''(x)}{n(x)} + \frac{1}{4} \frac{n'^2(x)}{n(x)} + \frac{2m}{\hbar^2} V(x) n(x) = \frac{2m}{\hbar^2} \mu n(x).$$

6.12. Consider the expression (6.201). Show that the functional derivative $\delta E_K / \delta n(\bar{x})$ is

$$\frac{\delta E_K}{\delta n(\bar{x})} = \frac{\hbar^2}{2m} \left(-\frac{1}{2} \frac{n''(\bar{x})}{n(\bar{x})} + \frac{1}{4} \frac{n'^2(\bar{x})}{n^2(\bar{x})} \right).$$

Show also that

$$\int n(x) \frac{\delta E_K}{\delta n(x)} dx = \frac{\hbar^2}{2m} \frac{1}{4} \int dx \frac{n'^2(x)}{n(x)} = E_K$$

Solution. Given $\delta n(x) = \delta n(\bar{x})\delta(x - \bar{x})$ and substituting in (6.201), one obtains $\delta E_K/\delta n(\bar{x})$. In addition,

$$\begin{aligned} \int n(x) \frac{\delta E_K}{\delta n(x)} dx &= \frac{\hbar^2}{2m} \left[-\frac{1}{2} \int n''(x) dx + \frac{1}{4} \int dx \frac{n'^2(x)}{n(x)} \right] \\ &= \frac{\hbar^2}{2m} \left[n'(x)|_{-\infty}^{\infty} + \frac{1}{4} \int dx \frac{n'^2(x)}{n(x)} \right] \\ &= \frac{\hbar^2}{2m} \frac{1}{4} \int dx \frac{n'^2(x)}{n(x)} = E_K \end{aligned} \quad (6.202)$$

References

1. J.C. Slater: Quantum theory of Matter, 5th ed. McGraw-Hill Book Company, New York (1985)
2. F. Bassani and U.M. Grassano: Fisica dello Stato Solido. 1st ed. Bollati Boringhieri, Torino (2000)
3. For a review, see e.g. W. Kohn, Rev. Mod. Phys. **71**, 1253 (1999)
4. For a review, see e.g. N.H. March, in *Theory of the inhomogeneous electron gas*. ed. by S. Lundqvist and N.H. March. Plenum Press, New York (1983); S. Lundqvist, ibid
5. M.L. Chiofalo, S. Conti, and M.P. Tosi, Mod. Phys. Lett. B **8**, 1207 (1994)
6. E. Kaxiras: Atomic and Electronic Structure of Solids. 1st ed. Cambridge Univ. Press (2003)
7. See e.g. P. Nozières and D. Pines: The Theory of Quantum Liquids - Superfluid Bose Liquids. Addison-Wesley, New York (1990)
8. In-Wan and E.W. Plummer, Phys. Rev. Lett. **60**, 1558 (1988)
9. R. Resta, Phys. Rev. B **16**, 2717 (1977)
10. D. Ninno, F. Trani, G. Cantele, K.J. Hameeuw, G. Iadonisi, E. Degoli, S. Ossicini, Europhys. Lett. **74**, 519 (2006)
11. For a pedagogical and compact theoretical understanding, see: G. Baym and C.J. Pethick, Phys. Rev. Lett. **76**, 6 (1996)
12. P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964)
13. W. Kohn and L.J. Sham, Phys. Rev. A **140**, 1133 (1965)
14. For a review available online, see e.g. P. Giannozzi: Density Functional Theory, <http://freescience.info/> (2002-2011)
15. N.D. Mermin, Phys. Rev. A **137**, 1441 (1965)
16. G. Vignale and W. Kohn, Phys. Rev. Lett. **77**, 2037 (1996)
17. For a review available online, see e.g. K. Burke and E.K.U. Gross: A guided tour of time-dependent density functional theory, <http://http://freescience.info/> (2002-2011)
18. G. Vignale, C. A. Ullrich, and S. Conti, Phys. Rev. Lett. **79**, 4878 (1997)
19. M. L. Chiofalo, A. Minguzzi, and M. P. Tosi, Physica B **254**, 188 (1998); ibid., Highlights INFM (1998-1999)
20. M. L. Chiofalo and M. P. Tosi, Europhys. Lett. **53**, 162 (2001)
21. R.O. Jones and O. Gunnarson, Rev. Mod. Phys. **61**, 689 (1989)
22. D.M. Ceperley, Phys. Rev. B **18**, 3126 (1978)
23. D.M. Ceperley and B.J. Alder, Phys. Rev. Lett. **45**, 566 (1980)

24. W. Hanke and L.J. Sham, Phys. Rev. B **12**, 4501 (1978)
25. L.J. Sham, in *Dynamical Properties of Solids*, ed. by G.H. Horton and A.A. Maradudin. North-Holland, Amsterdam (1974)
26. S. Ögüt, R. Burdick, Y. Saad, and J.R. Chelikowsky, Phys. Rev. Lett. **90**, 127401 (2003)
27. M. Hybertsen and S.G. Louie, Phys. Rev. B **35**, 5585 (1987)
28. F. Trani, D. Ninno, and G. Iadonisi, Phys. Rev. B **76**, 085326 (2006)
29. E.K.U. Gross and W. Kohn, Phys. Rev. Lett. **55**, 2850 (1985)
30. For a nice pedagogical illustration see, e.g. P. C. Martin: *Measurements and Correlation Functions* in Many-Body Physics, C. De Witt and R. Balian Eds. Gordon and Breach, New York (1968)

Subjects index

A

Absorption

- coefficient 5.2.1, 5.2.2.1, 5.2.3.1, 5.3.4, 5.4.3, 5.5.1
- transition probability 3.12, 3.18, 5.3, 5.4.3
- spectroscopy 2.2.8, 2.5.6, 2.8.2, 2.9.5, 3.12, 4.2.6.2, 5.2.2.4, 5.5.1, 5.6

Acoustical band 3.5.1.3, 3.7.2, 3.7.3

Acceptor states 4.4.4, 4.13

Al (Aluminum) 1.3.1, 1.6.7.3, 4.3.3.4, 4.3.5.8, 6.4.1, 6.5.3

Amorphous material 1.1, 1.8.2.1, 2.2.8

Analytic complex functions 5.2.2.1, 5.8

Anharmonic effects 3.10, 3.10.1, 3.10.2, 3.10.3, 3.11.1, 3.12.2, 4.6.5.3, 5.4.3.1

Atomic force microscopy (AFM) 1.6.7.3

Atomic structure factor 1.6.3, 1.6.4

Au (Gold) 1.2.3, 1.3, 1.6.7.3, 1.9, 2.2.7, 3.2.1.2, 4.3.3.4, 4.3.5.8

Auger spectroscopy 1.6.7.2, 2.9.7

B

Band 2.2.3, 2.2.6, 2.2.8, 2.3, 2.4, 2.5.5, 2.5.6, 2.12, 3.5.1, 3.6, 3.7, 3.17, 6.10.1

- width 2.2.3, 2.2.6, 2.2.7, 3.5.1.3, 3.7.2
- gap 2.2.6, 3.5.1.3
- offset 2.6.1

Base 1.2, 1.3, 1.6.3, 1.6.4, 2.3.4, 3.7

BCS 2.8.4, 3.11, 4.5

Bending 3.2.1.2

Bloch

- equation 4.2.6.2
- oscillations 4.2.1.1
- theorem 2.3, 2.4, 3.7, 3.15, 3.17
- wall 2.7.5
- Bond 1.3.1, 1.3.2, 1.6.4, 1.7, 1.8, 1.9, 2.2.8, 2.3.4, 2.9, 4.4.4, 6.7.1, 6.13
- Born-Oppenheimer method 2.5.1, 3.4, 3.11
- Boundary condition
 - atom 6.5.3, 6.13
 - nanocrystal 1.11, 2.2, 2.3.4, 3.5, 3.6, 3.14, 3.17
 - crystal 2.3, 2.4.1, 3.7
- Bravais lattice 1.2, 1.3, 1.6.3, 2.3, 2.4,
- Bragg law 1.6.5, 1.6.6, 2.4.1, 4.2.1
- Bremsstrahlung emission 1.6.7.2
- Brillouin
 - critical points 2.3, 3.7, 3.8.1
 - function 2.7.3, 2.7.4
 - scattering 3.12.4, 5.6
 - zone 1.4.2, 2.3, 2.4.1, 2.5.5, 2.6.1, 3.7, 4.2.1, 6.4.1

C

- C (Carbon) - nanotube 2.3.2, 2.3.4, 2.9.4, 2.9.6 - diamond 1.3.1.1, 1.6.4, 1.7.3, 1.7.4, 1.9, 2.4.3, 2.9.3, 4.6.5.3, 6.5.3, 6.10.2 - graphene 1.2.3, 2.3.2, 2.3.4, 6.10.1 - graphite 1.3.1.1, 1.7.4, 2.9.3, 5.2.2.3, 5.3.4.1 - C₆₀ 2.9.3
- Carbon nanotube 2.3.2, 2.3.4, 2.9.4, 2.9.6
- Carrier 2.3.3, 2.6, 2.8, 3.10.3, 4.2, 4.3, 4.4, 4.5.1, 4.6, 4.7, 4.8, 4.15, 5.2
- Chemical potential 2.5.3, 2.8.1, 2.13, 3.16, 6.5.3, 6.6.2, 6.6.3, 6.7.1, 6.11.2, 6.13
 - metal 4.3.0, 4.3.4, 5.4.1, 6.4.3
 - semiconductor 4.4, 4.7, 4.13
 - superconductor 4.5.1
 - nanostructure 4.8
- Cohesive energy 1.7, 3.4, 3.9, 3.11.1
- Compressibility 1.7.2, 1.7.3, 1.7.5, 3.2.1, 6.4.3, 6.11.2
- Compliance coefficient 3.2.1
- Correlation function 1.1, 1.9.1, 6.2.3, 6.5.2, 6.7.2, 6.7.3
- Core states 2.2, 2.4.3
- Covalent bond 1.3.1.1, 1.6.4, 1.7.4, 1.8, 1.9, 3.12.4
- Creation (destruction) operator
 - fermion 2.14.1, 3.11.2
 - boson 2.14.2, 3.8, 3.12.2, 3.12.4, 5.4.3
- Critical point
 - band 2.3.3, 3.8.1, 5.3.4

- superfluidity or superconductivity 2.8.3,
- Crystal
 - structure 1.1, 1.2, 1.3, 1.6, 1.7, 1.8, 1.9,
 - field 2.2.3
 - momentum 4.2, 6.6.3
- CsCl 1.3.2, 1.7.3, 1.9
- Cu (Copper) 1.9, 2.2.1, 2.5.5, 2.9.3, 2.9.4, 4.3.5.1, 4.3.5.4
- Current
 - electric 4.3.1
 - energy 4.3.2
 - density 2.8.3, 4.2.3, 5.7
 - drift 4.3.2, 4.3.3, 4.15
 - diffusion 4.15
- Current-Voltage characteristics
 - junction p-n 4.7.0.4
 - nanostructure 4.8.1, 4.8.3
- Cyclotron
 - frequency 4.2, 4.2.6.1, 4.3.3.4, 4.3.5.9, 6.5.3
 - mass 4.2.2, 4.9

D

- Debye
 - model 3.9, 4.3.5.2
 - temperature 3.9, 4.3.1, 4.6.2
- Deformation 1.9, 3.2, 3.3, 4.9
 - wave 3.2.1.2, 3.7.1, 3.17
- Density electronic states 2.3.4, 2.5.5, 2.9.4, 3.8.1, 4.3.4, 4.11, 5.3.4.1, 5.5.1
- Density vibrational mode 3.8.1
- Density Functional Theory (DFT) 6.6, 6.7, 6.8, 6.9, 6.10, 6.11
- Depletion 4.7.0.4, 4.15, 6.2
- Diamagnetism 2.7.2, 2.8.2
- Diamond 1.3.1.1, 1.6.4, 1.7.3, 1.7.4, 1.9, 2.4.3, 2.9.3, 4.6.5.3, 6.5.3, 6.10.2
- Dielectric constant 2.5.6, 6.10.3
 - static 4.4.4, 4.7.0.4, 5.2.2.4, 5.2.3.2, 5.2.3.3, 6.5.3, 6.10.3
 - high frequency 5.2.3.2, 5.2.3.3
- Dielectric function 5.2, 5.5, 5.6, 6.10.3
 - longitudinal 5.3.1, 5.3.3, 5.4, 5.7, 5.9, 6.10.2, 6.5.2, 6.5.3, 6.8, 6.10.2, 6.10.3, 6.13
 - transverse 5.3.2, 5.3.4, 5.7
 - sum rule 5.2.1, 5.3.3.2, 5.10
- Diffusion

- current 2.8.3, 4.15, 6.11.2
- length 4.15
- Dislocation 1.9,
- Distribution function 1.1, 2.13, 4.3.4, 4.4.1, 4.13
- Donor states 4.4.4, 4.13
- Ductility 1.8, 1.9

E

- Effective mass 2.2.3, 2.2.7, 2.3.3, 2.3.4, 2.5.5, 2.5.6, 2.6.1, 2.6.2, 2.7.6, 2.9, 4.2.2.1, 4.2.4, 4.2.6.1, 4.4.1, 4.4.4, 4.6.4, 4.6.5, 5.2.3.1, 6.10.3
- Elasticity 1.9, 3.2.1.1, 3.3
 - bending 3.2.1.2
 - compliance coefficient 3.2.1
 - stiffness coefficient 3.2.1
 - shape 3.2.1.1
- Elastic wave 3.2.1.2, 3.6.1
- Einstein
 - Bose-Einstein condensation 2.8.1
 - Bose-Einstein statistics 2.13
 - specific heat 3.9
 - relations 4.15
- Electric conductivity 1.8, 2.2.8, 4.2.1.1, 4.2.6.1, 4.8, 4.9, 5.2.2, 5.7
 - metal 4.3, 4.6.4, 5.2.2.2, 5.2.3.1, 5.7, 6.11
 - semiconductor 4.4.0, 4.6.5, 4.14
 - nanostructure 4.8
 - superconductivity 2.8, 4.5
- Electron
 - affinity 1.7.3, 2.5.2, 2.5.3, 6.3
 - ionization 1.7.3, 2.5.2, 2.5.3, 4.4.4, 6.3
 - crystal motion in electric and magnetic field 4.2
 - spectroscopy 1.6.7.2
 - spin resonance 2.9.9
 - tunneling 1.6.7.3, 4.5.1
- Electron-electron interaction 2.5.2, 3.11.2, 4.6.3, 5.4.0, 5.4.1, 6.7.2
- Electron-phonon interaction 3.11, 3.19, 4.5, 4.6.2, 5.4.3, 5.4.3, 5.10
- Electronegativity 1.7.3, 1.7.4, 1.8
- Electron energy
 - crystal 2.3, 2.4, 2.5.5, 6.3, 6.10.1
 - nanostructures 2.2, 2.6, 4.8, 6.2,
 - depending on density 6.4, 6.5, 6.6, 6.7
- Electromagnetic transition 5.3.4, 5.4.3.2, 5.5.1, 5.6

- electromagnetic induced by phonon 5.4.3.1
- Electron spectroscopy 1.6.7.2, 2.9
- Emission probability 3.18, 5.3.1, 5.3.2, 5.6
- Energy gap 2.2.6, 2.3.3, 2.4.1.2, 2.4.1.3, 2.5.5, 2.5.6, 2.6.1, 2.8.2, 3.5.1.3, 3.7.2, 4.4, 4.5.2, 6.10
- Energy Loss 2.9.2, 2.9.3, 2.9.4, 3.12.2, 5.2.3.1, 5.3.1, 5.3.3.2
- Exchange
 - energy 1.7.4, 1.7.5, 2.5, 2.7.4, 2.7.5, 2.11, 2.12, 6.2.1, 6.3, 6.4.4, 6.5.4, 6.6.2, 6.9.1
 - potential 2.7.4, 6.5.4
 - correlation 6.2, 6.4.4, 6.5.2, 6.5.5, 6.6.2, 6.7.1, 6.7.3, 6.8, 6.9, 6.10.2, 6.11
 - correlation hole 6.2.3, 6.7.2
 - correlation sum rule 6.2.3, 6.7.2
- Exciton 2.5.3,
 - crystal 2.5.6, 2.7.3, 2.7.6
 - nanometric system 2.6.2, 5.5.1
- Ewald construction 1.6.6

F

- Fermi
 - energy (wavelength) 1.11, 2.3.3, 2.8.4, 4.3.4, 4.3.5, 6.4.1
- Fermi-Dirac
 - distribution function 4.3.4
 - integrals 4.12
 - exchange energy 6.5.4
- Ferromagnetism (anti) 2.7.4, 2.7.5
- Folding 2.3.3
- Fresnel relations 5.2.2.3
- Form factor 2.4.3.1

G

- GaAs 1.3.2, 1.7.4, 2.5.5, 2.6.1, 3.6, 5.2.3.2, 5.5.1, 6.10.1
- Geometrical factor 1.6.3, 3.2.1.2
- Germanium 1.3.1.1, 1.7.4, 1.9, 3.7.3, 3.9, 5.3.4, 6.10.3
- Graphene 1.2.3, 2.3.2, 2.3.4, 6.10.1
- Graphite 1.3.1.1, 1.7.4, 2.9.3, 5.2.2.3, 5.3.4.1
- Gross-Kohn theorem 6.11.2

Group velocity

- electron 4.2.0, 4.2.1.1, 4.2.2
- electromagnetic wave 5.2.2.1
- phonon 3.7.1

H

Hall effect

- metals 4.3.3.4, 4.3.5.8
- quantum 4.3.5.9
- semiconductor 4.6.5.2, 4.14

Hardness 1.9

Harmonic approximation 2.5.1, 3.4, 3.9, 3.13

Hartree method 2.5.2, 6.4.3, 6.9.1

Hartree-Fock method 2.5.2, 2.5.4, 6.3, 6.4.4, 6.9.1

Hexagonal packed structure 1.2.2.2

Hg (mercury) 1.3.2, 2.8.2

Hohenberg-Kohn theorem 6.6.0, 6.6.1

- Mermin 6.6.1

Hole 2.5.3, 2.5.6, 2.6, 2.8.4.2, 4.2.5, 4.3.5.8, 4.4.1, 4.4.3, 4.4.4, 4.6.2, 4.6.5, 4.7.0.4, 4.13, 4.15, 5.5.1, 5.6, 6.2.3, 6.5.2, 6.5.4, 6.7.2, 6.8, 6.10.3

Hybridization 1.7.4, 2.3.4

I

Infrared spectroscopy 3.12.3, 3.12.4

Integrals improper 5.8.2

Ionic bonds 1.7

Ionicity 1.7.4,

J

Jellium 2.5.4, 2.12, 4.2.6.1, 6.4.1, 6.5.3

Josephson effect 4.5.1

Junction 4.5.1, 4.7, 4.8, 4.15

K

Kohn-Sham equations 6.6.2, 6.6.3, 6.7.2, 6.8, 6.11.2

Kramers-Kronig relations 5.2.2.1, 5.2.2.4, 5.3.1, 5.3.2, 5.3.3.2, 5.5

L

Landau

- levels 4.3.5.9, 4.9, 4.11

- hydrodynamical equation 6.11.2

Lattice 3.6.3,

- sites 1.2

- planes 1.5

- point group 1.2.1,

- reciprocal 1.4

- parameters 1.3

- symmetry 1.2.2

Lennard-Jones potential 1.7.2

Lindhard dielectric function 5.3.3.1, 5.4.1, 5.9, 6.9.1

Local density approximation (LDA) 6.5.4, 6.5.5, 6.5.6, 6.7.1, 6.11.2

Luminescence 1.6.7.2, 2.9.6, 5.6

Lyddane-Sachs-Teller relation 5.2.2.4, 5.2.3.2

M

Mean free path 2.2.8, 2.9.3, 4.3, 4.6.1, 4.6.3, 4.6.5.3, 4.15

Madelung constant 1.7.3, 1.7.5

Magnetization

- motion in magnetic field 4.2.2.2, 4.2.6.2

- susceptibility 2.7.1, 2.7.2, 2.7.3, 2.7.4, 2.7.5, 2.8.2, 3.3, 4.3.3.5, 4.3.5.7, 5.6

Magnetoconductivity 3.3.3.4

Malleability 1.11

Mass-action law 4.4.3

Matrix element

- constant operator 2.11

- single-particle operator 2.11

- two-particle operator 2.11

Metallic bond 1.7.5

Metallicity 1.8

Miller indexes 1.5.2
Mobility 4.6.5.1, 4.15

N

NaCl 1.3.2, 1.7.3, 3.2.1.2, 3.9, 5.2.3.2
Nanostructure 1.1, 1.2.3, 1.6.7, 2.2.1, 2.2.8, 2.3.4, 3.6, 3.12, 5.6
- bending 3.2.1.2
- electric conduction 4.8
- electronic energy 2.2.3, 2.2.4, 2.2.7, 2.5.3, 2.6, 2.9, 4.2.1.1, 6.10.3
- optical properties 5.5, 6.10.3
- specific heat 4.3.5.2
- vibrational frequency 3.5,
Navier-Stokes equations 6.11
Néel temperature 2.7.5
Noble gas 1.7.2
Normal mode 3.4
- quantization 3.8
Nuclear Magnetic Resonance (NMR) 4.2.6.2

O

Occupation number 2.5.3, 2.13, 2.14, 3.8, 3.9, 3.12.2, 3.12.4, 4.8.2, 5.4.3.1, 6.8
Optical band 3.5.1.3, 3.7.2
Overlap parameter 2.2.1, 2.2.2, 2.2.4, 2.2.5, 2.3.4, 2.7.4, 2.8.2, 4.8.2

P

Packing index 1.3.1, 1.7.5
Pair density 6.2
Paramagnetism 2.7.3
- Pauli paramagnetism 4.3.3.5
Parameter r_s 6.4.2
·Penetration depth 1.6.2, 2.8.3
·Plane wave (PW) method 2.4.1, 2.4.3,
·Phonon 3.8

- entropy, energy, free energy 3.16
- Photo-emission spectroscopy 2.9.3
- Piezoelectricity 3.3
- Piezomagnetism 3.3
- Plasticity 1.9
- Plasma frequency 5.2.1, 5.3.1, 5.6
 - electronic 5.2.3.1, 5.3.3.2
 - ionic 5.2.3.2, 5.4.2
- Plasmon 2.9.3, 5.2.3.1, 5.3.3.1
- Poisson
 - coefficient 3.2.1.1
 - equation 4.7.0.4, 5.4.1
- Polarization 1.7.2, 3.3, 6.13
 - susceptibility 5.2.1, 5.2.2, 5.2.3.3, 5.4.1, 5.7
- Polycrystal 1.1, 1.9
 - powder 1.6.6
- Power
 - absorbed 4.2.6.2,
 - dissipated 5.2.1, 5.3.1, 5.3.3.1
- Primitive cell 1.2.1, 1.2.2, 1.3.1
- Pseudopotential 2.4.3, 2.5.5, 6.4.1, 6.5.3, 6.10.1

Q

Quantum

- confinement 2.6, 5.5.1
- single-particle effect 2.6.1
- exciton states 2.6.2
- wells 2.6.1, 5.5.1
- wires 2.6.1
- dot 2.6.1, 2.6.2, 6.10.3
- electronic and optical properties 6.10.3

R

- Rabi frequency 3.18
- Raman spectroscopy 3.12.4, 5.6
- Rayleigh Scattering 2.9.4
- Reflectivity 5.2.2.3, 5.2.3.1, 5.2.3.3, 5.6

Refraction index 5.2.2.1, 5.2.2.3, 5.2.3.1
 Resilience 1.9
 Response function 5.2, 6.8
 Rigidity 3.2.1.1

S

Scattering in hydrodynamical regime 4.2.6, 5.2.1
 Scattering of
 - electron from impurities and phonons 2.9, 4.3, 4.6,
 - neutron from crystal 3.13.2
 - radiation from crystal and nanostructure 1.6, 2.9, 5.6
 Scanning electron microscopy (SEM) 1.6.7.2
 Scanning probe microscopy (SPM) 1.6.7.3
 Scanning tunnelling microscopy (STM) 1.6.7.3
 Screening 5.2.1, 5.4.0
 - electron-electron 5.4.1, 6.2.3, 6.5.2, 6.5.3, 6.7.2, 6.10.3, 6.13
 - ion-ion 5.4.2
 - effect in insulator 6.13
 Selection rules 1.6.2, 1.6.3, 3.12.2, 3.12.4, 3.19.1, 4.6.2, 4.9, 5.3.4.1, 5.4.3.1, 5.5.1, 5.6
 Semiconductor
 - band 2.3.3, 2.5.5, 4.2.1.1,
 - conductivities 4.2.6, 4.4, 4.6.3, 4.6.5, 4.7, 4.8, 4.9, 4.14
 - exciton 2.5.6
 - hole 4.2.5,
 - impurity states 4.4, 4.13,
 - optical properties 5.2.2.4, 5.3.3.2, 5.3.4.1, 5.4.3.1, 5.4.3.2, 5.5.1, 6.10.3
 Shear elasticity 1.9, 3.2.1, 6.11.1
 Si (silicon) 1.3.1.1, 1.6.7.2, 1.7.4, 1.9, 2.2.8, 2.3.2, 2.5.5, 3.7.3, 3.9, 4.3.5.1, 4.4.1, 6.5.3, 6.10
 Sommerfeld
 - model 4.3.4, 4.3.5, 4.6.3, 4.6.4, 5.2.3.1, 5.4.1
 - integrals 4.10
 Sound propagation 3.12.1
 Space group 1.2.2
 Specific heat
 - insulator 3.9, 3.10.1,
 - metal 4.3.3.2, 4.3.4, 4.3.5.1, 4.3.5.2
 - semiconductor 3.9, 4.3.5.1, 4.3.5.2
 - superconductor 2.8.2, 4.5.2,
 Stiffness coefficient 3.2.1

Stress 3.2.1

Stretching 3.2.1.1

Statistics

- Bose-Einstein 2.8, 2.13, 3.9, 6.5.3

- Fermi-Dirac 4.3.4, 4.13

Stokes (anti) line 3.12.4

Stress 1.9, 3.2, 3.3, 6.11.1

Superconductivity 2.8, 4.5

- critical temperature 2.8.1

- conductivity 4.5.1

- magnetic properties 4.5.3

Superfluidity 2.8.1, 2.8.3

Surface states 2.2.5

Symmetry operations 1.2.2, 2.3, 3.2.1, 3.12.4

T

Thermal conductivity 1.9, 4.9, 6.11

- insulators 3.10.3, 4.6.5.3

- metal 4.3.2, 4.3.4, 4.3.5.3, 4.6.4.2

Thermoelectric effect 4.3.2, 4.6.4.3

Thomas-Fermi

- model 5.4.1, 5.4.2, 6.5.1, 6.5.3, 6.9.2

- correlation energy 6.5.5

- Dirac model 6.5.4

- insulator screening effect 6.10.3, 6.13

- ultracold bosonic atoms 6.5.3

Tight-binding method (TB) 2.2.1, 2.2.2, 2.2.7, 2.3.4, 2.4.3

Time-Dependent Density Functional Theory (TD-DFT) 6.11

Tin 1.3.1.1, 1.7.4, 1.9

Transition probability 3.18, 5.3.1, 5.3.2

Transmission electron microscopy (TEM) 1.6.7.2

Transmittance 5.2.2.3

Two-fluid model 2.8.3, 6.11

U

Uemura plot 2.8.1,

Umklapp process 4.6.5.3, 5.4.3.2

V

Van der Waals 1.7.2

Variational calculus 2.10

Vector

- polarization 5.4.3.2,
- translation 1.2.1,
- reciprocal lattice 1.4.1, 1.4.2,

Vibrational mode

- nanocrystal 3.4, 3.5
- crystal 3.7, 3.17
- Vignale-Kohn 6.11.2
- Virial theorem 6.4.2, 6.5.3, 6.12,

X

X-ray spectroscopy 1.6, 2.9.3, 2.9.4, 2.9.5, 2.9.6

Y

Young modulus 3.2.1.1, 3.2.1.2

W

Wannier

- function 2.4.2
- exciton 2.5.6
- Stark state 4.2.1.1

Wave propagation 3.2.1.2, 3.7.1, 5.2.2.1, 5.2.3.3

Wiedemann and Franz law 4.3.3.3, 4.3.5.6

Wigner-Seitz cell 1.2.1

Z

ZnS 1.3.2, 1.7.3, 1.9

ZnSe 1.3.2, 1.7.4, 2.5.6,